(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 April 2002 (18.04.2002)

PCT

(10) International Publication Number
WO 02/31198 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/US01/31607

(22) International Filing Date: 11 October 2001 (11.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/239,294	11 October 2000 (11.10.2000)	US
60/239,297	11 October 2000 (11.10.2000)	US
60/239,605	11 October 2000 (11.10.2000)	US
60/239,802	12 October 2000 (12.10.2000)	US
60/239,805	12 October 2000 (12.10.2000)	US
60/239,806	12 October 2000 (12.10.2000)	US
60/240,622	16 October 2000 (16.10.2000)	US
60/241,723	19 October 2000 (19.10.2000)	US
60/241,682	19 October 2000 (19.10.2000)	US
60/244,932	31 October 2000 (31.10.2000)	US

(71) Applicant *(for all designated States except US)*: **AVALON PHARMACEUTICALS** [US/US]; 19 Firstfield Road, Gaithersburg, MD 20878 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **YOUNG, Paul, E.** [US/US]; 122 Beckwith Street, Gaithersburg, MD 20870 (US). **HORRIGAN, Stephen** [US/US]; 1895 Milboro Drive, Potomac, MD 20854 (US). **WEAVER, Zoe** [US/US]; 4615 North Park Avenue #1108, Chevy Chase, MD 20815 (US). **ENDRESS, Gregory, A.** [US/US]; 408 Bridge Road, Florence, MA 01062 (US).

(74) Agents: **GRANT, Alan, J.** et al.; Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road, Roseland, NJ 07068 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/31198 A2

(54) Title: CANCER-LINKED GENES AS TARGETS FOR CHEMOTHERAPY

(57) Abstract: Cancer-linked gene sequences, and derived amino acid sequences, are disclosed along with processes for assaying potential antitumor agents based on their modulation of the expression of these cancer-linked genes. Also disclosed are antibodies that react with the disclosed polypeptides and methods of diagnosing and treating cancer using the gene sequences. A novel gene and polypeptide are also disclosed.

CANCER-LINKED GENES AS TARGETS FOR CHEMOTHERAPY

This application claims the benefit of U.S. provisional application Serial No. 60/239,294, filed 11 October 2000; 60/239,297, filed 11 October 2000; 60/239,605, filed 11 October 2000; 60/239,802, filed 12 October 2000; 60/239,805, filed 12 October 2000; 60/239,806, filed 12 October 2000; 60/240,622, filed 16 October 2000; 60/241,682, filed 19 October 2000; 60/241,723, filed 19 October 2000; and 60/244,932, filed 31 October 2000, the disclosures of which are hereby incorporated by reference in their entirety.

FIELD OF THE INVENTION

The present invention relates to methods of screening cancer-linked genes and expression products for involvement in the cancer initiation and facilitation process and the use of such genes for screening potential anti-cancer agents, including small organic compounds and other molecules.

BACKGROUND OF THE INVENTION

Cancer-linked genes are valuable in that they indicate genetic differences between cancer cells and normal cells, such as where a gene is expressed in a cancer cell but not in a non-cancer cell, or where said gene is over-expressed or expressed at a higher level in a cancer as opposed to normal or non-cancer cell. In addition, the expression of such a gene in a normal cell but not in a cancer cell, especially of the same type of tissue, can indicate important functions in the cancerous process. For example, screening assays for novel drugs are based on the response of model cell based systems *in vitro* to treatment with specific compounds. Various measures of

cellular response have been utilized, including the release of cytokines, alterations in cell surface markers, activation of specific enzymes, as well as alterations in ion flux and/or pH. Some such screens rely on specific genes, such as oncogenes (or gene mutations). In accordance with the present invention, a cancer-linked gene has been identified and its putative amino acid sequence worked out. Such gene is useful in the diagnosing of cancer, the screening of anticancer agents and the treatment of cancer using such agents, especially in that these genes encode polypeptides that can act as markers, such as cell surface markers, thereby providing ready targets for anti-tumor agents such as antibodies, preferably antibodies complexed to cytotoxic agents, including apoptotic agents. .

BRIEF SUMMARY OF THE INVENTION

In accordance with the present invention, there is provided herein a set of genes related to, or linked to, cancer, or otherwise involved in the cancer initiating and facilitating process and the derived amino acid sequences thereof.

In a particular embodiment, such genes are those corresponding to the sequences of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19 and which encode polypeptides, including those comprising a sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20.

More particularly, such genes whose expression is changed in cancerous, as compared to non-cancerous cells, from a specific tissue, for example, lung, where the gene would include a polynucleotide corresponding to the nucleotide sequence of SEQ ID NO: 1 or sequences that are substantially identical to said sequence and/or encode the polypeptide with amino acid sequence of SEQ ID NO: 2 or a polypeptide differing therefrom by conservative amino acid substitutions..

It is another object of the present invention to provide methods of using such characteristic genes as a basis for assaying the potential ability of selected chemical agents to modulate upward or downward the expression of said cancer characteristic, or related, genes.

It is a further object of the present invention to provide methods of detecting the expression, or non-expression, or amount of expression, of said characteristic gene, or portions thereof, as a means of determining the cancerous, or non-cancerous, status (or potential cancerous status) of selected cells as grown in culture or as maintained *in situ*.

It is a still further object of the present invention to provide methods for treating cancerous conditions utilizing selected chemical agents as determined from their ability to modulate (i.e., increase or decrease) the characteristic gene, or its protein product.

The present invention also relates to a process for treating cancer comprising contacting a cancerous cell with an agent having activity against an expression product encoded by the genes, which process may be conducted either *ex vivo* or *in vivo* and which product is disclosed herein. Such agents may comprise an antibody or other molecule or portion that is specific for said expression product. In a preferred embodiment, the polypeptide product of such genes is a polypeptide as disclosed herein, such as SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20.

DETAILED SUMMARY OF THE INVENTION

The present invention relates to processes for utilizing a nucleotide sequence for a cancer-linked gene (SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19) and the derived amino acid sequence (SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20) as targets for chemotherapeutic agents, especially anti-cancer agents.

Characteristic gene sequences whose expression, or non-expression, or change in expression, are indicative of the cancerous or non-cancerous status of a given cell and whose expression is changed in cancerous, as compared to non-cancerous cells, from a specific tissue, are genes that include the nucleotide sequences disclosed herein or sequences that are substantially identical to said sequence, at least about 90% identical, preferably 95% identical, most preferably at least about 98% identical and especially where such gene has the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19. Such sequences have been searched within the GenBank database, with the following results.

The present invention relates to nucleotide sequences and derived polypeptides having the following characteristics:

A nucleotide sequence with Genbank Accession Numbers: NM_014109 and AA235448, representing an amplified bromodomain-containing protein in cancer that was identified with the following specific characteristics: **UniGene Cluster:** Hs. 46677; **Locus Link ID:** none; **Sequence Information:** 1934 bp mRNA (cDNA is SEQ ID NO: 1) 1086 bp ORF and 362 amino acids (SEQ ID NO: 2) with **Chromosomal Location:** 8q24 (based on alignment to mapped human genomic sequence). The deposited information represented a prediction of coding sequence deduced from a cDNA clone of unknown function. In accordance with the present invention, this message was up-regulated by at least 3-fold in lung cancer versus normal lung tissue. A search against the Prosite database reveals a domain with 100% similarity to the bromodomain 2 sequence, which is contained in a number of transcription factors. A key role for bromodomain proteins in maintaining normal proliferation is indicated by the implication of several bromodomain proteins in cancer, with four of these identified at translocation breakpoints.

A novel gene identified based upon EST sequences present within dbEST. This novel gene represents a novel member of the family of Toll-like

receptors and a portion of the polypeptide derived therefrom has at least 39% sequence identity to human toll-like receptor 1. Five of the human Toll-like receptors (called TLRs 1-5) may be direct homologs of the corresponding fly molecule and may constitute an important component of innate immunity in humans. Expression analysis shows that this gene is specifically expressed by B-lymphocytes. Characteristics were: **Genbank Accession Number for sample EST in cluster:** AA648836; **UniGene Cluster:** Hs.89206; **Locus Link ID:** 10330; **Cluster Name:** ESTs, Weakly similar to TLR6 [H.sapiens]; **Sequence Information:** 1274 bp mRNA (cDNA is SEQ ID NO: 3 with encoded polypeptide SEQ ID NO: 4. The UniGene cluster is composed of 10 sequences, all derived from tonsil. Microarray expression analysis indicates specific expression in B-lymphocytes.

A nucleotide sequence with Genbank Accession Number: AB015631 with the following specific characteristics: **Genbank Accession Number:** AB015631; **UniGene Cluster:** Hs.8752; **Locus Link ID:** 10330; **Cluster Name:** Transmembrane protein 4; **Sequence Information:** 814 bp mRNA (cDNA is SEQ ID NO: 5 with encoded polypeptide as SEQ ID NO: 6) and **Chromosomal Location:** 12. The UniGene cluster is composed of over 22 sequences derived from a number of tissues. Strongest levels of expression in normal tissues were detected in skeletal muscle.

A nucleotide sequence with Genbank Accession Number: NM_014397, AB026289 and with the following specific characteristics: **UniGene Cluster:** Hs.9625; **Locus Link ID:** 27073; **Public Cluster Name:** SID6-1512, putative serine-threonine protein kinase; **Sequence Information:** 1597 bp mRNA (SEQ ID NO: 7), 921 bp ORF and 307 amino acids (and SEQ ID NO: 8) with **Chromosomal Location:** 9q33. From the record in Genbank the complete mRNA and protein sequences are obtained. SID6-1512 shares sequence similarity with murine NEK1, a kinase involved in cell cycle regulation. The UniGene cluster contains over 150 EST sequences from a variety of tissue sources. The top BLAST score was to protein kinase nek1, which contains an N-terminal protein kinase domain with about 42% identity to the catalytic domain of NIMA, a protein kinase that controls initiation of mitosis in

Aspergillus nidulans. In addition, both Nek1 and NIMA have a long, basic C-terminal extension and are therefore similar in overall structure.

A nucleotide sequence with Genbank Accession Number: NM_006035, AF128625 and with the following specific characteristics: **UniGene Cluster:** Hs.12908; **Locus Link ID:** 9578; **Cluster Name:** CDC42-binding protein kinase beta (DMPK-like; MRCKbeta); **Sequence Information:** 6780 bp mRNA (SEQ ID NO: 9), 5136 bp ORF and 1711 amino acids (SEQ ID NO: 10) with **Chromosomal Location:** 14q32.3. The UniGene cluster contained over 215 EST sequences from a variety of tissue sources. The p21 GTPases, Rho and Cdc42, regulate numerous cellular functions by binding to members of a serine/threonine protein kinase subfamily. These functions include the remodeling of the cell cytoskeleton that is a feature of cell growth and differentiation. Two of these p21 GTPase-regulated kinases, the myotonic dystrophy protein kinase-related Cdc42-binding kinases (MRCKalpha and beta), have been demonstrated to phosphorylate nonmuscle myosin light chain, a prerequisite for the activation of actin-myosin contractility. A BLAST search showed A portion of SEQ ID NO: 10 to have about 49% identity to human myotonic dystrophy kinase.

A nucleotide sequence with Genbank Accession Number: NM_002654 with the following specific characteristics: **UniGene Cluster:** Hs.198281; **Locus Link ID:** 5315; **Cluster Name:** Pyruvate kinase, muscle; **Sequence Information:** 2287 bp mRNA (SEQ ID NO: 11 with derived amino acid sequence SEQ ID NO: 12)) with **Chromosomal Location:** 15q22. This gene in a member of a small sub-family within the cdk family of protein kinases. PCTIARE-3 appears to play a role in signal transduction in terminally differentiated cells. The cloning of the human and murine PCTAIRE-3 genes have been described but no other information is available in the scientific literature. The UniGene cluster is composed of 64 sequences derived from a number of tissues. PCTAIRE-3 was expressed in colon adenocarcinomas tested and was expressed at a lower level or not at all in normal colon tissue samples tested.

A nucleotide sequence with Genbank Accession Number: NM_006293 with the following specific characteristics: **UniGene Cluster:** Hs.301; **Locus Link ID:** 7301; **Cluster Name:** TYRO3 protein tyrosine kinase; **Sequence Information:** 4364 bp mRNA (SEQ ID NO: 13 with derived amino acid sequence SEQ ID NO: 14) with **Chromosomal Location:** 15q15.1-q21.1. The UniGene cluster is composed of over 45 sequences derived from a number of tissues. Strongest levels of expression in normal tissues detected in brain. SEQ ID NO: 14 displays appreciable homology to a variety of receptor tyrosine kinases. For example, a portion of SEQ ID NO: 14 has at least about 43% identity to AXL receptor tyrosine kinase. Over-expression of axl cDNA in NIH 3T3 cells induces neoplastic transformation with the concomitant appearance of a 140 kD axl tyrosine-phosphorylated protein.

A nucleotide sequence with Genbank Accession Number: NM_002969 with the following specific characteristics: **UniGene Cluster:** Hs.55039; **Locus Link ID:** 6300; **Cluster Name:** Mitogen-activated protein kinase 12; **Sequence Information:** 1457 bp mRNA (SEQ ID NO: 15 with derived amino acid sequence of SEQ ID NO: 16) with **Chromosomal Location:** 22q13.33. The UniGene cluster is composed of over 22 sequences derived from a number of tissues. The strongest levels of expression in normal tissues were detected in skeletal muscle. This sequence displays appreciable homology to a variety of mitogen-activated protein kinases, for example, with human mitogen-activated protein kinase p38delta. The p38 mitogen-activated protein kinases (MAPK) play a crucial role in stress and inflammatory responses and are also involved in activation of the human immunodeficiency virus gene expression.

A nucleotide sequence with Genbank Accession Number: W31344 with the following specific characteristics: **UniGene Cluster:** Hs.55444; **Locus Link ID:** unknown; **Cluster Name:** ESTs; **Chromosomal Location:** unknown. The UniGene cluster is composed of over 9 sequences, all of which are derived from parathyroid. The GenBank database shows an exact match to AF153819 (*Homo sapiens* inwardly-rectifying potassium channel Kir2.1). This match is entirely confined to the 3' untranslated region of the GenBank entry.

The translation product of thyrocarcin and the Kir2.1 gene are identical; however, we cannot formally rule out the possibility that thyrocarcin is a completely different gene that shares some splicing with Kir2.1. SEQ ID NO: 17 shows the nucleotide sequence for Kir2.1 and SEQ ID NO: 21 shows EST cluster identified from expression analysis that is specific for thyroid adenocarcinoma. The derived amino acid sequence from SEQ ID NO: 17 is shown as SEQ ID NO: 18. The sequence of the EST cluster displays no obvious homology to known proteins. However, for the case where the EST cluster is simply the 3'-untranslated region of Kir2.1, this gene is an inwardly-rectifying potassium channel.

A nucleotide sequence with Genbank Accession Number: AA133334, representing a Sox2-like HMG-box Oncogenically Expressed Sequence with the following specific characteristics: **UniGene Cluster:** Hs.129911; **Locus Link ID:** none; **Sequence Information:** 1050 bp mRNA (bp = base pair, SEQ ID NO: 19), 264 bp ORF and 88 amino acids (SEQ ID NO: 20) with **Chromosomal Location:** unknown. SEQ ID NO: 19 is present as an EST (Expressed Sequence Tag) in the Gene Logic database. It has been elongated to 1050 bp by overlapping contigs in the public databases. The unigene cluster indicates widespread expression and it was found that the message is upregulated by at least 3-fold in lung cancer versus normal lung tissue. This sequence significant homolgy with Ovis aries SOX-2 gene, and to a slightly lesser extent the murine SOX-2. The Sox gene family consists of a large number of embryonically expressed genes related via the possession of a 79-amino-acid DNA-binding domain known as the HMG box. These genes are transcription factors likely to be involved in the regulation of gene expression.

The nucleotides and polypeptides, as gene products, used in the processes of the present invention may comprise a recombinant polynucleotide or polypeptide, a natural polynucleotide or polypeptide, or a synthetic polynucleotide or polypeptide, preferably a recombinant polynucleotide or polypeptide.

Fragments of such polynucleotides and polypeptides as are disclosed herein may also be useful in practicing the processes of the present invention. For example, a fragment, derivative or analog of the polypeptide (SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20) may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

In one aspect, the present invention relates to an isolated polynucleotide comprising a polynucleotide at least 65% identical to the polynucleotide of SEQ ID NO: 3, or its complement. In preferred embodiments, said isolated polynucleotide comprises a polynucleotide that has sequence identity of at least 80%, preferably at least about 90%, most preferably at least about 95%, especially at least about 98% and most especially is identical to the sequence of SEQ ID NO: 3. An isolated polynucleotide of the invention may also include the complement of any of the foregoing.

In another aspect, the present invention relates to an isolated polypeptide, including a purified polypeptide, comprising an amino acid sequence at least 90% identical to the amino acid sequence of SEQ ID NO: 4. In preferred embodiments, said isolated polypeptide comprises an amino acid sequence having sequence identity of at least 95%, preferably at least about 98%, and especially is identical to, the sequence of SEQ ID NO: 4. The present invention also includes isolated active fragments of such polypeptides where said fragments retain the biological activity of the polypeptide or where such

active fragments are useful as specific targets for cancer treatment, prevention or diagnosis.

The polynucleotides and polypeptides useful in practicing the processes of the present invention may likewise be obtained in an isolated or purified form. In addition, the polypeptide disclosed herein as being useful in practicing the processes of the invention include different types of proteins in terms of function so that, as recited elsewhere herein, some are enzymes, some are transcription factors and other may be cell surface receptors. Precisely how such cancer-linked proteins are used in the processes of the invention may thus differ depending on the function and cellular location of the protein and therefore modification, or optimization, of the methods disclosed herein may be desirable in light of said differences. For example, a cell-surface receptor is an excellent target for cytotoxic antibodies whereas a transcription factor or enzyme is a useful target for a small organic compound with anti-neoplastic activity.

As used herein, the term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). It could also be produced recombinantly and subsequently purified. For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides, for example, those prepared recombinantly, could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment. In one embodiment of the present invention, such isolated, or purified, polypeptide is useful in generating antibodies for practicing the invention, or where said antibody is attached to a cytotoxic or cytolytic agent, such as an apoptotic agent.

As known in the art "similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide.

The sequence information disclosed herein, as derived from the GenBank submissions, can readily be utilized by those skilled in the art to prepare the corresponding full-length polypeptide by peptide synthesis. The same is true for either the polynucleotides or polypeptides disclosed herein for use in the methods of the invention.

As used herein, the terms "portion," "segment," and "fragment," when used in relation to polypeptides, refer to a continuous sequence of residues, such as amino acid residues, which sequence forms a subset of a larger sequence. For example, if a polypeptide were subjected to treatment with any of the common endopeptidases, such as trypsin or chymotrypsin, the oligopeptides resulting from such treatment would represent portions, segments or fragments of the starting polypeptide. When used in relation to a polynucleotides, such terms refer to the products produced by treatment of said polynucleotides with any of the common endonucleases.

The present invention further relates to a vector comprising any of the polynucleotides disclosed herein and to a recombinant cell comprising such vectors, or such polynucleotides or expressing the polypeptides disclosed herein, especially the polypeptide whose amino acid sequence is the sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20.

Methods of producing such cells and vectors are well known to those skilled in the molecular biology art. See, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), Wu et al, *Methods in Gene Biotechnology* (CRC Press, New York, NY, 1997), and *Recombinant Gene Expression Protocols*, in *Methods in Molecular Biology*, Vol. 62, (Tuan, ed., Humana Press, Totowa, NJ, 1997), the disclosures of which are hereby incorporated by reference.

In another aspect, the present invention relates to a process for identifying an agent that modulates the activity of a cancer-related gene comprising:

(a) contacting a compound with a cell containing a gene that corresponds to a polynucleotide having a sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19 and under conditions promoting the expression of said gene; and

(b) detecting a difference in expression of said gene relative to when said compound is not present

thereby identifying an agent that modulates the activity of a cancer-related gene.

In specific embodiments of the present invention, the genes useful for the invention comprise genes that correspond to polynucleotides having a sequence selected from SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19, or may comprise the sequence of any of the polynucleotides disclosed herein (where the latter are cDNA sequences). As used herein, "corresponding genes" refers to genes that encode an RNA that is at least 90% identical, preferably at least 95% identical, most preferably at least 98% identical, and especially identical, to an RNA encoded by one of the nucleotide sequences disclosed herein (i.e., SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19). Such genes will also encode the same polypeptide sequence as any of the sequences disclosed herein, preferably SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20, but may include differences in such amino acid sequences where such differences are limited to conservative amino acid substitutions, such as where the same overall three dimensional structure, and thus the same antigenic character, is maintained. Thus, amino acid sequences may be within the scope of the present invention where they react with the same antibodies that react with polypeptides comprising the sequences of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20 as disclosed herein.

As used herein, the term "conservative amino acid substitution" are defined herein as exchanges within one of the following five groups:

I. Small aliphatic, nonpolar or slightly polar residues:

Ala, Ser, Thr, Pro, Gly;

II. Polar, negatively charged residues and their amides:

Asp, Asn, Glu, Gln;

III. Polar, positively charged residues:

His, Arg, Lys;

IV. Large, aliphatic, nonpolar residues:

Met Leu, Ile, Val, Cys

V. Large, aromatic residues:

Phe, Tyr, Trp

In accordance with the present invention, model cellular systems using cell lines, primary cells, or tissue samples are maintained in growth medium and may be treated with compounds that may be at a single concentration or at a range of concentrations. At specific times after treatment, cellular RNAs are isolated from the treated cells, primary cells or tumors, which RNAs are indicative of expression of selected genes. The cellular RNA is then divided and subjected to analysis that detects the presence and/or quantity of specific RNA transcripts, which transcripts may then be amplified for detection purposes using standard methodologies, such as, for example, reverse transcriptase polymerase chain reaction (RT-PCR), etc. The presence or absence, or levels, of specific RNA transcripts are determined from these measurements and a metric derived for the type and degree of response of the sample to the treated compound compared to control samples.

In accordance with the foregoing, there is thus disclosed herein processes for using a cancer-linked gene sequence (SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19) whose expression is, or can be, as a result of the methods of the present invention, linked to, or used to characterize, the cancerous, or non-cancerous, status of the cells, or tissues, to be tested. Thus, the processes of the present invention identify novel anti-neoplastic agents based on their alteration of expression of the polynucleotide sequence disclosed herein in specific model systems. The methods of the invention may therefore be used with a variety of cell lines or with primary samples from tumors maintained *in vitro* under suitable culture conditions for varying periods of time, or *in situ* in suitable animal models.

More particularly, genes have been identified that is expressed at a level in cancer cells that is different from the expression level in non-cancer cells. In one instance, the identified genes are expressed at higher levels in cancer cells than in normal cells.

The polynucleotides of the invention can include fully operation genes with attendant control or regulatory sequences or merely a polynucleotide sequence encoding the corresponding polypeptide or an active fragment or analog thereof.

In one embodiment of the present invention, said gene modulation is downward modulation, so that, as a result of exposure to the chemical agent to be tested, one or more genes of the cancerous cell will be expressed at a lower level (or not expressed at all) when exposed to the agent as compared to the expression when not exposed to the agent. For example, the gene encoding the polypeptide of SEQ ID NO: 2 is expressed at a higher level in cells of lung cancer than in normal lung cells.

In a preferred embodiment a selected set of said genes are expressed in the reference cell, including the gene(s) sequences identified for use according to the present invention, but are not expressed in the cell to be tested as a result of the exposure of the cell to be tested to the chemical agent. Thus, where said chemical agent causes the gene, or genes, of the tested cell to be expressed at a lower level than the same genes of the reference, this is indicative of downward modulation and indicates that the chemical agent to be tested has anti-neoplastic activity.

Sequences encoding the same proteins as any of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20, regardless of the percent identity of such sequences, are also specifically contemplated by any of the methods of the present invention that rely on any or all of said sequences, regardless of how they are otherwise described or limited. Thus, any such sequences are available for use in carrying out any of the methods disclosed according to the

invention. Such sequences also include any open reading frames, as defined herein, present within the sequence of SEQ ID NO: 1.

The genes identified by the present disclosure are considered "cancer-related" genes, as this term is used herein, and include genes expressed at higher levels (due, for example, to elevated rates of expression, elevated extent of expression or increased copy number) in cancer cells relative to expression of these genes in normal (i.e., non-cancerous) cells where said cancerous state or status of test cells or tissues has been determined by methods known in the art, such as by reverse transcriptase polymerase chain reaction (RT-PCR) as described in the Example below. In specific embodiments, this relates to the genes whose sequences correspond to the sequences of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19. As used herein, the term "correspond" means that the gene has the indicated nucleotide sequence or that it encodes substantially the same RNA as would be encoded by the indicated sequence, the term "substantially" meaning about at least 90% identical as defined elsewhere herein and includes splice variants thereof.

The sequences disclosed herein may be genomic in nature and thus represent the sequence of an actual gene, such as a human gene, or may be a cDNA sequence derived from a messenger RNA (mRNA) and thus represent contiguous exonic sequences derived from a corresponding genomic sequence or they may be wholly synthetic in origin for purposes of practicing the processes of the invention. Because of the processing that may take place in transforming the initial RNA transcript into the final mRNA, the sequences disclosed herein may represent less than the full genomic sequence. They may also represent sequences derived from ribosomal and transfer RNAs. Consequently, the genes present in the cell (and representing the genomic sequences) and the sequences disclosed herein, which are mostly cDNA sequences, may be identical or may be such that the cDNAs contain less than the full genomic sequence. Such genes and cDNA sequences are still considered corresponding sequences because they both encode similar RNA sequences. Thus, by way of non-limiting example only, a

gene that encodes an RNA transcript, which is then processed into a shorter mRNA, is deemed to encode both such RNAs and therefore encodes an RNA complementary to (using the usual Watson-Crick complementarity rules), or that would otherwise be encoded by, a cDNA (for example, a sequence as disclosed herein). Thus, the sequences disclosed herein correspond to genes contained in the cancerous or normal cells used to determine relative levels of expression because they represent the same sequences or are complementary to RNAs encoded by these genes. Such genes also include different alleles and splice variants that may occur in the cells used in the processes of the invention.

The genes of the invention "correspond to" a polynucleotide having a sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19 if the gene encodes an RNA (processed or unprocessed, including naturally occurring splice variants and alleles) that is at least 90% identical, preferably at least 95% identical, most preferably at least 98% identical to, and especially identical to, an RNA that would be encoded by, or be complementary to, such as by hybridization with, a polynucleotide having the indicated sequence. In addition, genes including sequences at least 90% identical to a sequence selected from SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19, preferably at least about 95% identical to such a sequence, more preferably at least about 98% identical to such sequence and most preferably comprising such sequence are specifically contemplated by all of the processes of the present invention as being genes that correspond to these sequences. In addition, sequences encoding the same proteins as any of these sequences, regardless of the percent identity of such sequences, are also specifically contemplated by any of the methods of the present invention that rely on any or all of said sequences, regardless of how they are otherwise described or limited. Thus, any such sequences are available for use in carrying out any of the methods disclosed according to the invention. Such sequences also include any open reading frames, as defined herein, present within any of the sequences of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19.

Further in accordance with the present invention, the term "percent identity" or "percent identical," when referring to a sequence, means that a sequence is compared to a claimed or described sequence after alignment of the sequence to be compared (the "Compared Sequence") with the described or claimed sequence (the "Reference Sequence"). The Percent Identity is then determined according to the following formula:

$$\text{Percent Identity} = 100\,[1-(C/R)]$$

wherein C is the number of differences between the Reference Sequence and the Compared Sequence over the length of alignment between the Reference Sequence and the Compared Sequence wherein (i) each base or amino acid in the Reference Sequence that does not have a corresponding aligned base or amino acid in the Compared Sequence and (ii) each gap in the Reference Sequence and (iii) each aligned base or amino acid in the Reference Sequence that is different from an aligned base or amino acid in the Compared Sequence, constitutes a difference; and R is the number of bases or amino acids in the Reference Sequence over the length of the alignment with the Compared Sequence with any gap created in the Reference Sequence also being counted as a base or amino acid.

If an alignment exists between the Compared Sequence and the Reference Sequence for which the percent identity as calculated above is about equal to or greater than a specified minimum Percent Identity then the Compared Sequence has the specified minimum percent identity to the Reference Sequence even though alignments may exist in which the hereinabove calculated Percent Identity is less than the specified Percent Identity.

As used herein and except as noted otherwise, all terms are defined as given below.

In accordance with the present invention, the term "DNA segment" or "DNA sequence" refers to a DNA polymer, in the form of a separate fragment

or as a component of a larger DNA construct, which has been derived from DNA isolated at least once in substantially pure form, i.e., free of contaminating endogenous materials and in a quantity or concentration enabling identification, manipulation, and recovery of the segment and its component nucleotide sequences by standard biochemical methods, for example, using a cloning vector. Such segments are provided in the form of an open reading frame uninterrupted by internal nontranslated sequences, or introns, which are typically present in eukaryotic genes. Sequences of non-translated DNA may be present downstream from the open reading frame, where the same do not interfere with manipulation or expression of the coding regions.

The term "coding region" refers to that portion of a gene which either naturally or normally codes for the expression product of that gene in its natural genomic environment, i.e., the region coding *in vivo* for the native expression product of the gene. The coding region can be from a normal, mutated or altered gene, or can even be from a DNA sequence, or gene, wholly synthesized in the laboratory using methods well known to those of skill in the art of DNA synthesis.

In accordance with the present invention, the term "nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the proteins provided by this invention are assembled from cDNA fragments and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

The term "expression product" means that polypeptide or protein that is the natural translation product of the gene and any nucleic acid sequence coding equivalents resulting from genetic code degeneracy and thus coding for the same amino acid(s).

The term "active fragment," when referring to a coding sequence, means a portion comprising less than the complete coding region whose expression

product retains essentially the same biological function or activity as the expression product of the complete coding region.

The term "primer" means a short nucleic acid sequence that is paired with one strand of DNA and provides a free 3'-OH end at which a DNA polymerase starts synthesis of a deoxyribonucleotide chain.

The term "promoter" means a region of DNA involved in binding of RNA polymerase to initiate transcription. The term "enhancer" refers to a region of DNA that, when present and active, has the effect of increasing expression of a different DNA sequence that is being expressed, thereby increasing the amount of expression product formed from said different DNA sequence.

The term "open reading frame (ORF)" means a series of triplets coding for amino acids without any termination codons and is a sequence (potentially) translatable into protein.

As used herein, reference to a DNA sequence includes both single stranded and double stranded DNA. Thus, the specific sequence, unless the context indicates otherwise, refers to the single strand DNA of such sequence, the duplex of such sequence with its complement (double stranded DNA) and the complement of such sequence.

The present invention also relates to methods of assaying potential antitumor agents based on their modulation of the expression of the gene sequence according to the invention and methods for diagnosing cancerous, or potentially cancerous, conditions as a result of the patterns of expression of the gene sequence disclosed herein as well as related gene sequence based on common expression or regulation of such genes.

In carrying out the foregoing assays, relative antineoplastic activity may be ascertained by the extent to which a given chemical agent modulates the expression of genes present in a cancerous cell. Thus, a first chemical agent that modulates the expression of a gene associated with the cancerous state

(i.e., a gene that includes one of the sequences of the invention as disclosed herein and present in cancerous cells) to a larger degree than a second chemical agent tested by the assays of the invention is thereby deemed to have higher, or more desirable, or more advantageous, anti-neoplastic activity than said second chemical agent.

The gene expression to be measured is commonly assayed using RNA expression as an indicator. Thus, the greater the level of RNA (messenger RNA) detected the higher the level of expression of the corresponding gene. Thus, gene expression, either absolute or relative, is determined by the relative expression of the RNAs encoded by such genes.

RNA may be isolated from samples in a variety of ways, including lysis and denaturation with a phenolic solution containing a chaotropic agent (e.g., triazol) followed by isopropanol precipitation, ethanol wash, and resuspension in aqueous solution; or lysis and denaturation followed by isolation on solid support, such as a Qiagen resin and reconstitution in aqueous solution; or lysis and denaturation in non-phenolic, aqueous solutions followed by enzymatic conversion of RNA to DNA template copies.

Normally, prior to applying the processes of the invention, steady state RNA expression levels for the genes, and sets of genes, disclosed herein will have been obtained. It is the steady state level of such expression that is affected by potential anti-neoplastic agents as determined herein. Such steady state levels of expression are easily determined by any methods that are sensitive, specific and accurate. Such methods include, but are in no way limited to, real time quantitative polymerase chain reaction (PCR), for example, using a Perkin-Elmer 7700 sequence detection system with gene specific primer probe combinations as designed using any of several commercially available software packages, such as Primer Express software., solid support based hybridization array technology using appropriate internal controls for quantitation, including filter, bead, or microchip based arrays, solid support based hybridization arrays using, for example, chemiluminescent, fluorescent, or electrochemical reaction based detection systems.

The gene patterns indicative of a cancerous state need not be characteristic of every cell found to be cancerous. Thus, the methods disclosed herein are useful for detecting the presence of a cancerous condition within a tissue where less than all cells exhibit the complete pattern. Thus, for example, a set of selected genes, comprising sequences corresponding to the sequence of SEQ ID NO: 1, may be found, using appropriate probes, either DNA or RNA, to be present in as little as 60% of cells derived from a sample of tumorous, or malignant, tissue while being absent from as much as 60% of cells derived from corresponding non-cancerous, or otherwise normal, tissue (and thus being present in as much as 40% of such normal tissue cells). In a preferred embodiment, such gene pattern is found to be present in at least 50% of cells drawn from a cancerous tissue, such as the lung cancer disclosed herein. In an additional embodiment, such gene pattern is found to be present in at least 100% of cells drawn from a cancerous tissue and absent from at least 100% of a corresponding normal, non-cancerous, tissue sample, although the latter embodiment may represent a rare occurrence.

In another aspect the present invention relates to a process for determining the cancerous status of a test cell, comprising determining expression in said test cell of a gene sequence as disclosed herein and then comparing said expression to expression of said at least one gene in at least one cell known to be non-cancerous whereby a difference in said expression indicates that said cell is cancerous.

In one embodiment, said change in expression is a change in copy number, including either an increase or decrease in copy number. In accordance with the present invention, said change in gene copy number may be determined by determining a change in expression of messenger RNA encoded by said gene sequence.

Changes in gene copy number may be determined by determining a change in expression of messenger RNA encoded by a particular gene

sequence, especially that of Such change in gene copy number may be determined by determining a change in expression of messenger RNA encoded by a particular gene sequence, especially that of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19. Also in accordance with the present invention, said gene may be a cancer initiating gene, a cancer facilitating gene, or a cancer suppressing gene. In carrying out the methods of the present invention, a cancer facilitating gene is a gene that, while not directly initiating or suppressing tumor formation or growth, said gene acts, such as through the actions of its expression product, to direct, enhance, or otherwise facilitate the progress of the cancerous condition, including where such gene acts against genes, or gene expression products, that would otherwise have the effect of decreasing tumor formation and/or growth.

Although the presence or absence of expression of a gene corresponding to a sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19 may be indicative of a cancerous status for a given cell, the mere presence or absence of such a gene may not alone be sufficient to achieve a malignant condition and thus the level of expression of such gene pattern may also be a significant factor in determining the attainment of a cancerous state. Thus, while a pattern of genes may be present in both cancerous and non-cancerous cells, the level of expression, as determined by any of the methods disclosed herein, all of which are well known in the art, may differ between the cancerous versus the non-cancerous cells. Thus, it becomes essential to also determine the level of expression of a gene such as that disclosed herein, including substantially similar sequences and sequences comprising said sequence, as a separate means of diagnosing the presence of a cancerous status for a given cell, groups of cells, or tissues, either in culture or *in situ*.

The level of expression of the polypeptides disclosed herein is also a measure of gene expression, such as polypeptides having sequence identical, or similar to any polypeptide encoded by the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19, such as the polypeptide whose amino acid sequence is the sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20.

In accordance with the foregoing, the present invention further relates to a process for determining the cancerous status of a cell to be tested, comprising determining the level of expression in said cell of at least one gene that includes one of the nucleotide sequences selected from the sequences of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19, including sequences substantially identical to said sequences, or characteristic fragments thereof, or the complements of any of the foregoing and then comparing said expression to that of a cell known to be non-cancerous whereby the difference in said expression indicates that said cell to be tested is cancerous.

In accordance with the invention, although gene expression for a gene that includes as a portion thereof one of the sequences of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19, is preferably determined by use of a probe that is a fragment of such nucleotide sequence, it is to be understood that the probe may be formed from a different portion of the gene. Expression of the gene may be determined by use of a nucleotide probe that hybridizes to messenger RNA (mRNA) transcribed from a portion of the gene other than the specific nucleotide sequence disclosed herein.

It should be noted that there are a variety of different contexts in which genes have been evaluated as being involved in the cancerous process. Thus, some genes may be oncogenes and encode proteins that are directly involved in the cancerous process and thereby promote the occurrence of cancer in an animal. In addition, other genes may serve to suppress the cancerous state in a given cell or cell type and thereby work against a cancerous condition forming in an animal. Other genes may simply be involved either directly or indirectly in the cancerous process or condition and may serve in an ancillary capacity with respect to the cancerous state. All such types of genes are deemed with those to be determined in accordance with the invention as disclosed herein. Thus, the gene determined by said process of the invention may be an oncogene, or the gene determined by said process may be a cancer facilitating gene, the latter including a gene that directly or indirectly affects the cancerous process, either in the promotion of a cancerous condition or in facilitating the progress of cancerous growth or

otherwise modulating the growth of cancer cells, either *in vivo* or *ex vivo*. In addition, the gene determined by said process may be a cancer suppressor gene, which gene works either directly or indirectly to suppress the initiation or progress of a cancerous condition. Such genes may work indirectly where their expression alters the activity of some other gene or gene expression product that is itself directly involved in initiating or facilitating the progress of a cancerous condition. For example, a gene that encodes a polypeptide, either wild or mutant in type, which polypeptide acts to suppress of tumor suppressor gene, or its expression product, will thereby act indirectly to promote tumor growth.

In accordance with the foregoing, the process of the present invention includes cancer modulating agents that are themselves either polypeptides, or small chemical entities, that affect the cancerous process, including initiation, suppression or facilitation of tumor growth, either *in vivo* or *ex vivo*. Said cancer modulating agent may have the effect of increasing gene expression or said cancer modulating agent may have the effect of decreasing gene expression as such terms have been described herein.

In keeping with the disclosure herein, the present invention also relates to a process for treating cancer comprising contacting a cancerous cell with an agent having activity against an expression product encoded by a gene sequence as disclosed herein, such as the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19.

The proteins encoded by the genes disclosed herein due to their expression, or elevated expression, in cancer cells, represent highly useful therapeutic targets for "targeted therapies" utilizing such affinity structures as, for example, antibodies coupled to some cytotoxic agent. In such methodology, it is advantageous that nothing need be known about the endogenous ligands or binding partners for such cell surface molecules. Rather, an antibody or equivalent molecule that can specifically recognize the cell surface molecule (which could include an artificial peptide, a surrogate ligand, and the like) that is coupled to some agent that can induce cell death

or a block in cell cycling offers therapeutic promise against these proteins. Thus, such approaches include the use of so-called suicide "bullets" against intracellular proteins

With the advent of methods of molecular biology and recombinant technology, it is now possible to produce antibody molecules by recombinant means and thereby generate gene sequences that code for specific amino acid sequences found in the polypeptide structure of the antibodies. Such antibodies can be produced by either cloning the gene sequences encoding the polypeptide chains of said antibodies or by direct synthesis of said polypeptide chains, with *in vitro* assembly of the synthesized chains to form active tetrameric (H_2L_2) structures with affinity for specific epitopes and antigenic determinants. This has permitted the ready production of antibodies having sequences characteristic of neutralizing antibodies from different species and sources.

Regardless of the source of the antibodies, or how they are recombinantly constructed, or how they are synthesized, *in vitro* or *in vivo*, using transgenic animals, such as cows, goats and sheep, using large cell cultures of laboratory or commercial size, in bioreactors or by direct chemical synthesis employing no living organisms at any stage of the process, all antibodies have a similar overall 3 dimensional structure. This structure is often given as H_2L_2 and refers to the fact that antibodies commonly comprise 2 light (L) amino acid chains and 2 heavy (H) amino acid chains. Both chains have regions capable of interacting with a structurally complementary antigenic target. The regions interacting with the target are referred to as "variable" or "V" regions and are characterized by differences in amino acid sequence from antibodies of different antigenic specificity.

The variable regions of either H or L chains contains the amino acid sequences capable of specifically binding to antigenic targets. Within

these sequences are smaller sequences dubbed "hypervariable" because of their extreme variability between antibodies of differing specificity. Such hypervariable regions are also referred to as "complementarity determining regions" or "CDR" regions. These CDR regions account for the basic specificity of the antibody for a particular antigenic determinant structure.

The CDRs represent non-contiguous stretches of amino acids within the variable regions but, regardless of species, the positional locations of these critical amino acid sequences within the variable heavy and light chain regions have been found to have similar locations within the amino acid sequences of the variable chains. The variable heavy and light chains of all antibodies each have 3 CDR regions, each non-contiguous with the others (termed L1, L2, L3, H1, H2, H3) for the respective light (L) and heavy (H) chains. The accepted CDR regions have been described by Kabat et al, *J. Biol. Chem.* **252**:6609-6616 (1977). The numbering scheme is shown in the figures, where the CDRs are underlined and the numbers follow the Kabat scheme.

In all mammalian species, antibody polypeptides contain constant (i.e., highly conserved) and variable regions, and, within the latter, there are the CDRs and the so-called "framework regions" made up of amino acid sequences within the variable region of the heavy or light chain but outside the CDRs.

The antibodies disclosed according to the invention may also be wholly synthetic, wherein the polypeptide chains of the antibodies are synthesized and, possibly, optimized for binding to the polypeptides disclosed herein as being receptors. Such antibodies may be chimeric or humanized antibodies and may be fully tetrameric in structure, or may be dimeric and comprise only a single heavy and a single light chain. Such antibodies may also include fragments, such as Fab and $F(ab_2)'$ fragments, capable of reacting with and binding to any of the polypeptides disclosed herein as being receptors.

In one aspect, the present invention relates to immunoglobulins, or antibodies, as described herein, that react with, especially where they are specific for, the polypeptides having amino acid sequences as disclosed herein, preferably those having an amino acid sequence of one of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20. Such antibodies may commonly be in the form of a composition, especially a pharmaceutical composition.

The pharmaceutical compositions useful herein also contain a pharmaceutically acceptable carrier, including any suitable diluent or excipient, which includes any pharmaceutical agent that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Pharmaceutically acceptable carriers include, but are not limited to, liquids such as water, saline, glycerol and ethanol, and the like, including carriers useful in forming sprays for nasal and other respiratory tract delivery or for delivery to the ophthalmic system. A thorough discussion of pharmaceutically acceptable carriers, diluents, and other excipients is presented in REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. current edition).

The process of the present invention includes embodiments of the above-recited processes wherein the cancer cell is contacted *in vivo* as well as *ex vivo*, preferably wherein said agent comprises a portion, or is part of an overall molecular structure, having affinity for said expression product. In one such embodiment, said portion having affinity for said expression product is an antibody, especially where said expression product is a polypeptide or oligopeptide or comprises an oligopeptide portion, or comprises a polypeptide.

Such an agent can therefore be a single molecular structure, comprising both affinity portion and anti-cancer activity portions, wherein said portions are derived from separate molecules, or molecular structures,

possessing such activity when separated and wherein such agent has been formed by combining said portions into one larger molecular structure, such as where said portions are combined into the form of an adduct. Said anti-cancer and affinity portions may be joined covalently, such as in the form of a single polypeptide, or polypeptide-like, structure or may be joined non-covalently, such as by hydrophobic or electrostatic interactions, such structures having been formed by means well known in the chemical arts. Alternatively, the anti-cancer and affinity portions may be formed from separate domains of a single molecule that exhibits, as part of the same chemical structure, more than one activity wherein one of the activities is against cancer cells, or tumor formation or growth, and the other activity is affinity for an expression product produced by expression of genes related to the cancerous process or condition.

In one embodiment of the present invention, a chemical agent, such as a protein or other polypeptide, is joined to an agent, such as an antibody, having affinity for an expression product of a cancerous cell, such as a polypeptide or protein encoded by a gene related to the cancerous process, especially a gene as disclosed herein according to the present invention. Thus, where the presence of said expression product is essential to tumor initiation and/or growth, binding of said agent to said expression product will have the effect of negating said tumor promoting activity. In one such embodiment, said agent is an apoptosis-inducing agent that induces cell suicide, thereby killing the cancer cell and halting tumor growth..

Other genes within the cancer cell that are regulated in a manner similar to that of the genes disclosed herein and thus change their expression in a coordinated way in response to chemical compounds represent genes that are located within a common metabolic, signaling, physiological, or functional pathway so that by analyzing and identifying such commonly regulated groups of genes (groups that include the gene, or similar sequences, disclosed according to the invention, one can (a) assign known genes and novel genes to specific pathways and (b) identify specific functions and functional roles for novel genes that are grouped into pathways with

genes for which their functions are already characterized or described. For example, one might identify a group of 10 genes, at least one of which is the gene as disclosed herein, that change expression in a coordinated fashion and for which the function of one, such as the polypeptide encoded by the sequence disclosed herein, is known then the other genes are thereby implicated in a similar function or pathway and may thus play a role in the cancer-initiating or cancer-facilitating process. In the same way, if a gene were found in normal cells but not in cancer cells, or happens to be expressed at a higher level in normal as opposed to cancer cells, then a similar conclusion may be drawn as to its involvement in cancer, or other diseases. Therefore, the processes disclosed according to the present invention at once provide a novel means of assigning function to genes, i.e. a novel method of functional genomics, and a means for identifying chemical compounds that have potential therapeutic effects on specific cellular pathways. Such chemical compounds may have therapeutic relevance to a variety of diseases outside of cancer as well, in cases where such diseases are known or are demonstrated to involve the specific cellular pathway that is affected.

The polypeptides disclosed herein, preferably those of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20, also find use as vaccines in that, where the polypeptide represents a surface protein present on a cancer cell, such polypeptide may be administered to an animal, especially a human being, for purposes of activating cytotoxic T lymphocytes (CTLs) that will be specific for, and act to lyze, cancer cells in said animal. Where used as vaccines, such polypeptides are present in the form of a pharmaceutical composition. The present invention may also employ polypeptides that have the same, or similar, immunogenic character as the polypeptides of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20 and thereby elicit the same, or similar, immunogenic response after administration to an animal, such as an animal at risk of developing cancer, or afflicted therewith. Thus, the polypeptides disclosed according to the invention will commonly find use as immunogenic compositions.

The present invention also relates to a process that comprises a method for producing a product comprising identifying an agent according to one of the disclosed processes for identifying such an agent (i.e., the therapeutic agents identified according to the assay procedures disclosed herein) wherein said product is the data collected with respect to said agent as a result of said identification process, or assay, and wherein said data is sufficient to convey the chemical character and/or structure and/or properties of said agent. For example, the present invention specifically contemplates a situation whereby a user of an assay of the invention may use the assay to screen for compounds having the desired enzyme modulating activity and, having identified the compound, then conveys that information (i.e., information as to structure, dosage, etc) to another user who then utilizes the information to reproduce the agent and administer it for therapeutic or research purposes according to the invention. For example, the user of the assay (user 1) may screen a number of test compounds without knowing the structure or identity of the compounds (such as where a number of code numbers are used the first user is simply given samples labeled with said code numbers) and, after performing the screening process, using one or more assay processes of the present invention, then imparts to a second user (user 2), verbally or in writing or some equivalent fashion, sufficient information to identify the compounds having a particular modulating activity (for example, the code number with the corresponding results). This transmission of information from user 1 to user 2 is specifically contemplated by the present invention.

The genes useful in the methods of the invention disclosed herein are genes corresponding to a polynucleotide having the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 or 19 and represent genes that may be over-expressed in malignant cancer, such as is a gene corresponding to SEQ ID NO: 1 for lung, the latter being expressed at least three fold higher in lung cancer samples as compared to normal lung tissues. In addition, in any given sample, not all cancer cells may express this gene a substantial expression thereof in a substantial number of such cells is sufficient to warrant a determination of a cancerous, or potentially cancerous, condition.

Thus, the polynucleotide sequences disclosed according to the present invention are expressed in cancer compared to normal tissue samples or may be expressed at a higher level in cancer as compared to normal tissues. Further, such polynucleotide, or gene, sequence expression in normal tissues may correlate with individuals having a family history of cancer.

Such gene sequences may play a direct role in cancer progression, such as in cancer initiation or cancer cell proliferation/survival. For example, one or more genes encoding the same polypeptide as one or more of the sequences disclosed herein represent novel individual gene targets for screening and discovery of small molecules that inhibit enzyme or other cellular functions, e.g. kinase inhibitors. Such molecules represent valuable therapeutics for cancer. In addition, small molecules or agents, such as small organic molecules, that down-regulate the expression of these genes in cancer would represent valuable anti-cancer therapeutics. Expression of the gene in normal tissues may indicate a predisposition towards development of lung cancer. The encoded polypeptide might represent a potentially useful cell surface target for therapeutic molecules such as cytolytic antibodies, or antibodies attached to cytotoxic, or cytolytic, agents. .

It should be cautioned that, in carrying out the procedures of the present invention as disclosed herein, any reference to particular buffers, media, reagents, cells, culture conditions and the like are not intended to be limiting, but are to be read so as to include all related materials that one of ordinary skill in the art would recognize as being of interest or value in the particular context in which that discussion is presented. For example, it is often possible to substitute one buffer system or culture medium for another and still achieve similar, if not identical, results. Those of skill in the art will have sufficient knowledge of such systems and methodologies so as to be able, without undue experimentation, to make such substitutions as will optimally serve their purposes in using the methods and procedures disclosed herein.

The present invention will now be further described by way of the following non-limiting example. In applying the disclosure of the example, it should be kept clearly in mind that other and different embodiments of the methods disclosed according to the present invention will no doubt suggest themselves to those of skill in the relevant art. The following example shows how a potential anti-neoplastic agent may be identified using one or more of the genes disclosed herein.

EXAMPLE

SW480 cells are grown to a density of 10^5 cells/cm^2 in Leibovitz's L-15 medium supplemented with 2 mM L-glutamine (90%) and 10% fetal bovine serum. The cells are collected after treatment with 0.25% trypsin, 0.02% EDTA at 37°C for 2 to 5 minutes. The trypsinized cells are then diluted with 30 ml growth medium and plated at a density of 50,000 cells per well in a 96 well plate (200 µl/well). The following day, cells are treated with either compound buffer alone, or compound buffer containing a chemical agent to be tested, for 24 hours. The media is then removed, the cells lysed and the RNA recovered using the RNAeasy reagents and protocol obtained from Qiagen. RNA is quantitated and 10 ng of sample in 1 µl are added to 24 µl of Taqman reaction mix containing 1X PCR buffer, RNAsin, reverse transcriptase, nucleoside triphosphates, amplitaq gold, tween 20, glycerol, bovine serum albumin (BSA) and specific PCR primers and probes for a reference gene (18S RNA) and a test gene (Gene X). Reverse transcription is then carried out at 48°C for 30 minutes. The sample is then applied to a Perlin Elmer 7700 sequence detector and heat denatured for 10 minutes at 95°C. Amplification is performed through 40 cycles using 15 seconds annealing at 60°C followed by a 60 second extension at 72°C and 30 second denaturation at 95°C. Data files are then captured and the data analyzed with the appropriate baseline windows and thresholds.

The quantitative difference between the target and reference gene is then calculated and a relative expression value determined for all of the samples used. This procedure is then repeated for other genes functionally related to the gene as disclosed herein and the level of function, or expression, noted. The relative expression ratios for each pair of genes is determined (i.e., a ratio of expression is determined for each target gene versus each of the other genes for which expression is measured, where each gene's absolute expression is determined relative to the reference gene for each compound, or chemical agent, to be screened). The samples are then scored and ranked according to the degree of alteration of the expression profile in the treated samples relative to the control. The overall expression of the particular gene relative to the controls, as modulated by one chemical agent relative to another, is also ascertained. Chemical agents having the most effect on a given gene, or set of genes, are considered the most anti-neoplastic.

WHAT IS CLAIMED IS:

1. A process for identifying an agent that modulates the activity of a cancer-related gene comprising:

(a) contacting a compound with a cell containing a gene that corresponds to a polynucleotide having a sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19 and under conditions promoting the expression of said gene; and

(b) detecting a difference in expression of said gene relative to when said compound is not present

thereby identifying an agent that modulates the activity of a cancer-related gene.

2. The process of claim 1 wherein said gene has a sequence selected from the group consisting of SEQ ID NO: SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19.

3. The process of claim 1 wherein the cell is a cancer cell and the difference in expression is a decrease in expression.

4. The process of claim 2 wherein the cell is a cancer cell and the difference in expression is a decrease in expression.

5. A process for identifying an anti-neoplastic agent comprising contacting a cell exhibiting neoplastic activity with a compound first identified as a cancer related gene modulator using a process of one of claims 1 – 4 and detecting a decrease in said neoplastic activity after said contacting compared to when said contacting does not occur.

6. The process of claim 5 wherein said neoplastic activity is accelerated cellular replication.

7. The process of claim 5 wherein said decrease in neoplastic activity results from the death of the cell.

8. A process for identifying an anti-neoplastic agent comprising administering to an animal exhibiting a cancer condition an effective amount of an agent first identified according to a process of one of claims 1-7 and detecting a decrease in said cancerous condition.

9. A process for determining the cancerous status of a cell, comprising determining an increase in the level of expression in said cell of at least one gene that corresponds to a polynucleotide having a sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19 wherein an elevated expression relative to a known non-cancerous cell indicates a cancerous state or potentially cancerous state.

10. An antibody that reacts with a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20.

11. The antibody of claim 10 wherein said antibody is a monoclonal antibody.

12. The antibody of claim 10 wherein said antibody is a recombinant antibody.

13. The antibody of claim 10 wherein said antibody is a synthetic antibody.

14. The antibody of claim 10 wherein said antibody further comprises a cytotoxic agent.

15. The antibody of claim 14 wherein said cytotoxic agent is an apoptotic agent.

16. A process for treating cancer comprising contacting a cancerous cell with an agent having activity against an expression product encoded by a

gene sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17 and 19.

17. The process of claim 16 wherein said cancerous cell is contacted *in vivo*.

18. The process of claim 16 wherein said agent has affinity for said expression product.

19. The process of claim 18 wherein said agent is an antibody of claim 10 – 15.

20. An immunogenic composition comprising a polypeptide comprising an amino acid sequence with at least 90% identity to a sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20 and wherein and amino acid difference results only from conservative amino acid substitutions.

21. The immunogenic composition of claim 20 wherein said percent identity is at least 95%.

22. The immunogenic composition of claim 20 wherein said percent identity is at least 98%.

23. The immunogenic composition of claim 20 wherein said polypeptide has the sequence of a member selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20.

24. A process for treating cancer in an animal afflicted therewith comprising administering to said animal an amount of an immunogenic composition of claim 20-23 sufficient to elicit the production of cytotoxic T lymphocytes specific for said immunogenic composition.

25. The process of claim 24 wherein said animal is a human being.

26. A process for treating a cancerous condition in an animal afflicted therewith comprising administering to said animal a therapeutically effective amount of an agent first identified as having anti-neoplastic activity using the process of claim 8.

27. A process for protecting an animal against cancer comprising administering to an animal at risk of developing cancer a therapeutically effective amount of an agent first identified as having anti-neoplastic activity using the process of claim 8.

28. A method for producing a product comprising identifying an agent according to the process of claim 1 - 8 wherein said product is the data collected with respect to said agent as a result of said process and wherein said data is sufficient to convey the chemical structure and/or properties of said agent.

29. An isolated polynucleotide comprising a polynucleotide having at least 95% sequence identity to a member selected from the group consisting of SEQ ID NO: 3 or the complement thereof.

30. The isolated polynucleotide of claim 29 wherein said polynucleotide comprises the sequence of SEQ ID NO: 3.

31. An isolated polynucleotide comprising a polynucleotide selected from the group consisting of:

(a) a polynucleotide encoding the amino acid sequence of SEQ ID NO: 4, and

(b) the complement of (a).

32. An isolated polynucleotide comprising an amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO: 4 and wherein any difference in sequence identity results only from conservative amino acid substitutions.

SEQUENCE LISTING

<110> Young, Paul
Horrigan, Stephen
Weaver, Zoe
Endress, Gregory

<120> Cancer-Linked Genes as Targets for Chemotherapy

<130> 689290-79

<150> US/60/239,294
<151> 2000-10-11

<150> US/60/239,297
<151> 2000-10-11

<150> US/60/239,605
<151> 2000-10-11

<150> US/60/239,802
<151> 2000-10-12

<150> US/60/239,805
<151> 2000-10-12

<150> US/60/239,806
<151> 2000-10-12

<150> US/60/240,622
<151> 2000-10-16

<150> US/60/241,682
<151> 2000-10-19

<150> US/60/241,723
<151> 2000-10-19

<150> US/60/244,932
<151> 2000-10-31

<160> 21

<170> PatentIn version 3.0

<210> 1
<211> 1934
<212> DNA
<213> Homo sapiens

<400> 1
ctgtcattca tgctttggaa aagtttactg tatatacatt agacattcct gttctttttg 60
gagttagtac tacatcccct gaagaaacat gtgcccaggt gattcgtgaa gctaagagaa 120
cagcaccaag tatagtgtat gttcctcata tccacgtgtg gtgggaaata gttggaccga 180
cacttaaagc cacatttacc acattattac agaatattcc ttcatttgct ccagttttac 240
tacttgcaac ttctgacaaa ccccattccg ctttgccaga agaggtgcaa gaattgttta 300
tccgtgatta tggagagatt tttaatgtcc agttaccgga taaagaagaa cggacaaaat 360
tttttgaaga tttaattcta aaacaagctg ctaagcctcc tatatcaaaa aagaaagcag 420

ttttgcaggc tttggaggta ctcccagtag caccaccacc tgagccaaga tcactgacag 480
cagaagaagt gaaacgacta gaagaacaag aagaagatac atttagagaa ctgaggattt 540
tcttaagaaa tgttacacat aggcttgcta ttgacaagcg attccgagtg tttactaagc 600
ctgttgaccc tgatgaggtt cctgattatg tcactgtaat aaagcaacca atggaccttt 660
catctgtaat cagtaaaatt gatctacaca agtatctgac tgtgaaagac tatttgagag 720
atattgatct aatctgtagt aatgccttag aatacaatcc agatagagat cctggagatc 780
gtcttattag gcatagagcc tgtgctttaa gagatactgc ctatgccata attaaagaag 840
aacttgatga agactttgag cagctctgtg aagaaattca ggaatctaga aagaaaagag 900
gttgtagctc ctccaaatat gccccgtctt actaccatgt gatgccaaag caaaattcca 960
ctcttgttgg tgataaaaga tcagacccag agcagaatga aaagctaaag acaccgagta 1020
ctcctgtggc ttgcagcact cctgctcagt tgaagaggaa aattcgcaaa aagtcaaact 1080
ggtacttagg caccataaaa aagcgaagga agatttcaca ggcaaaggat gatagccaga 1140
atgccataga tcacaaaatt gagagtgata cagaggaaac tcaagacaca agtgtagatc 1200
ataatgagac cggaaacaca ggagagtctt cggtggaaga aaatgaaaaa cagcaaaatg 1260
cctctgaaag caaactggaa ttgagaaata attcaaatac ttgtaatata gagaatgagc 1320
ttgaagactc taggaagact acagcatgta cagaattgag agacaagatt gcttgtaatg 1380
gagatgcttc tagctctcag ataatacata tttctgatga aaatgaagga aaagaaatgt 1440
gtgttctgcg aatgactcga gctagacgtt cccaggtaga acagcagcag ctcatcactg 1500
ttgaaaaggc tttggcaatt ctttctcagc ctacaccctc acttgttgtg gatcatgagc 1560
gattaaaaaa tcttttgaag actgttgtta aaaaaagtca aaactacaac atatttcagt 1620
tggaaaattt gtatgcagta atcagccaat gtatttatcg gcatcgcaag gaccatgata 1680
aaacatcact tattcagaaa atggagcaag aggtagaaaa cttcagttgt tccagatgat 1740
gatgtcatgg tatcgagtat tctttatatt cagttcctat ttaagtcatt tttgtcatgt 1800
ccgcctaatt gatgtagtat gaaaccctgc atctttaagg aaaagattaa aatagtaaaa 1860
taaaagtatt taaactttcc tgatatttat gtacatatta agataaatgt catgtgtaag 1920
ataactgata aata 1934

<210> 2
<211> 362
<212> PRT
<213> Homo sapiens

<400> 2
Met Asp Leu Ser Ser Val Ile Ser Lys Ile Asp Leu His Lys Tyr Leu
1 5 10 15

Thr Val Lys Asp Tyr Leu Arg Asp Ile Asp Leu Ile Cys Ser Asn Ala
20 25 30

Leu Glu Tyr Asn Pro Asp Arg Asp Pro Gly Asp Arg Leu Ile Arg His
35 40 45

Arg Ala Cys Ala Leu Arg Asp Thr Ala Tyr Ala Ile Ile Lys Glu Glu
50 55 60

Leu Asp Glu Asp Phe Glu Gln Leu Cys Glu Glu Ile Gln Glu Ser Arg
65 70 75 80

Lys Lys Arg Gly Cys Ser Ser Ser Lys Tyr Ala Pro Ser Tyr Tyr His
85 90 95

Val Met Pro Lys Gln Asn Ser Thr Leu Val Gly Asp Lys Arg Ser Asp
100 105 110

Pro Glu Gln Asn Glu Lys Leu Lys Thr Pro Ser Thr Pro Val Ala Cys
115 120 125

Ser Thr Pro Ala Gln Leu Lys Arg Lys Ile Arg Lys Lys Ser Asn Trp

130 135 140

Tyr Leu Gly Thr Ile Lys Lys Arg Arg Lys Ile Ser Gln Ala Lys Asp
145 150 155 160

Asp Ser Gln Asn Ala Ile Asp His Lys Ile Glu Ser Asp Thr Glu Glu
165 170 175

Thr Gln Asp Thr Ser Val Asp His Asn Glu Thr Gly Asn Thr Gly Glu
180 185 190

Ser Ser Val Glu Glu Asn Glu Lys Gln Gln Asn Ala Ser Glu Ser Lys
195 200 205

Leu Glu Leu Arg Asn Asn Ser Asn Thr Cys Asn Ile Glu Asn Glu Leu
210 215 220

Glu Asp Ser Arg Lys Thr Thr Ala Cys Thr Glu Leu Arg Asp Lys Ile
225 230 235 240

Ala Cys Asn Gly Asp Ala Ser Ser Ser Gln Ile Ile His Ile Ser Asp
245 250 255

Glu Asn Glu Gly Lys Glu Met Cys Val Leu Arg Met Thr Arg Ala Arg
260 265 270

Arg Ser Gln Val Glu Gln Gln Gln Leu Ile Thr Val Glu Lys Ala Leu
275 280 285

Ala Ile Leu Ser Gln Pro Thr Pro Ser Leu Val Val Asp His Glu Arg
290 295 300

Leu Lys Asn Leu Leu Lys Thr Val Val Lys Lys Ser Gln Asn Tyr Asn
305 310 315 320

Ile Phe Gln Leu Glu Asn Leu Tyr Ala Val Ile Ser Gln Cys Ile Tyr
325 330 335

Arg His Arg Lys Asp His Asp Lys Thr Ser Leu Ile Gln Lys Met Glu
340 345 350

Gln Glu Val Glu Asn Phe Ser Cys Ser Arg
355 360

<210> 3
<211> 1274
<212> DNA
<213> Homo sapiens

<400> 3
cacgaggcag actgaagatt gtggcttggt attcacaggc aggtttcaga catttagatc 60
tttcttttaa tgactaacac catgcctatc tgtggagaag ctggcaacat gtcacacctg 120
gaaattgttt ttcaacatta atactattat ttggcagtaa tccagattgc ttttgccacc 180
aacctgaaga catatagagg cagaaggaca ggaataattc tatttgtttc ctgttttgaa 240
acttccatct gtaaggctat caaaaggaga tgtgagagag ggtattgagt ctggcctgac 300
aatgcagttc ttaaaccaaa ggtccattat gcttctcctc tctgagaatc ctgacttacc 360
tcaacaacgg agacatggca cagtagccag cttggagact tctcagccaa tgctctgaga 420
tcaagtcgaa gacccaatat acagggtttt gagctcatct tcatcattca tatgaggaaa 480
taagtggtaa aatccttgga aatacaatga gactcatcag aaacatttac atattttgta 540

```
gtattgttat gacagcagag ggtgatgctc cagagctgcc agaagaaagg gaactgatga    600
ccaactgctc caacatgtct ctaagaaagg ttcccgcaga cttgacccca gccacaacga    660
cactggattt atcctataac ctccttttc aactccagag ttcagatttt cattctgtct    720
ccaaactgag agttttgatt ctatgccata acagaattca acagctggat ctcaaaacct    780
ttgaattcaa caaggagtta agatatttag atttgtctaa taacagactg aagagtgtaa    840
cttggtattt actggcaggt ctcaggtatt tagatctttc ttttaatgac tttgacacca    900
tgcctatctg tgaggaagct ggcaacatgt cacacctgga aatcctaggt ttgagtgggg    960
caaaaataca aaaatcagat ttccagaaaa ttgctcatct gcatctaaat actgtcttct   1020
taggattcag aactcttcct cattatgaag aaggtagcct gcccatctta aacacaacaa   1080
aactgcacat tgttttacca atggacacaa atttctgggt tcttttgcgt gatggaatca   1140
agacttcaaa aatattagaa atgacaaata tagatggcaa aagccaattt gtaagttatg   1200
aaatgcaacg aaatcttagt ttagaacatg ctaagacatc ggttctattg cttaataaag   1260
ttgatttact ctgg                                                     1274
```

<210> 4
<211> 256
<212> PRT
<213> Homo sapiens

<400> 4

```
Met Arg Leu Ile Arg Asn Ile Tyr Ile Phe Cys Ser Ile Val Met Thr
1               5                  10                  15

Ala Glu Gly Asp Ala Pro Glu Leu Pro Glu Glu Arg Glu Leu Met Thr
            20                  25                  30

Asn Cys Ser Asn Met Ser Leu Arg Lys Val Pro Ala Asp Leu Thr Pro
        35                  40                  45

Ala Thr Thr Thr Leu Asp Leu Ser Tyr Asn Leu Leu Phe Gln Leu Gln
    50                  55                  60

Ser Ser Asp Phe His Ser Val Ser Lys Leu Arg Val Leu Ile Leu Cys
65                  70                  75                  80

His Asn Arg Ile Gln Gln Leu Asp Leu Lys Thr Phe Glu Phe Asn Lys
                85                  90                  95

Glu Leu Arg Tyr Leu Asp Leu Ser Asn Asn Arg Leu Lys Ser Val Thr
            100                 105                 110

Trp Tyr Leu Leu Ala Gly Leu Arg Tyr Leu Asp Leu Ser Phe Asn Asp
        115                 120                 125

Phe Asp Thr Met Pro Ile Cys Glu Glu Ala Gly Asn Met Ser His Leu
    130                 135                 140

Glu Ile Leu Gly Leu Ser Gly Ala Lys Ile Gln Lys Ser Asp Phe Gln
145                 150                 155                 160

Lys Ile Ala His Leu His Leu Asn Thr Val Phe Leu Gly Phe Arg Thr
                165                 170                 175

Leu Pro His Tyr Glu Glu Gly Ser Leu Pro Ile Leu Asn Thr Thr Lys
            180                 185                 190

Leu His Ile Val Leu Pro Met Asp Thr Asn Phe Trp Val Leu Leu Arg
        195                 200                 205
```

Asp Gly Ile Lys Thr Ser Lys Ile Leu Glu Met Thr Asn Ile Asp Gly
210 215 220

Lys Ser Gln Phe Val Ser Tyr Glu Met Gln Arg Asn Leu Ser Leu Glu
225 230 235 240

His Ala Lys Thr Ser Val Leu Leu Leu Asn Lys Val Asp Leu Leu Trp
245 250 255

<210> 5
<211> 814
<212> DNA
<213> Homo sapiens

<400> 5
agaatcccgg acagccctgc tccctgcagc caggtgtagt ttcgggagcc actggggcca 60
aagtgagagt ccagcggtct tccagcgctt gggccacggc ggcggccctg ggagcagagg 120
tggagcgacc ccattacgct aaagatgaaa ggctggggtt ggctggccct gcttctgggg 180
gccctgctgg gaaccgcctg ggctcggagg agccaggatc tccactgtgg agcatgcagg 240
gctctggtgg atgaactaga atgggaaatt gcccaggtgg accccaagaa gaccattcag 300
atgggatctt tccggatcaa tccagatggc agccagtcag tggtggaggt gccttatgcc 360
cgctcagagg cccacctcac agagctgctg gaggagatat gtgaccggat gaaggagtat 420
ggggaacaga ttgatccttc cacccatcgc aagaactacg tacgtgtagt gggccggaat 480
ggagaatcca gtgaactgga cctacaaggc atccgaatcg actcagatat tagcggcacc 540
ctcaagtttg cgtgtgagag cattgtggag gaatacgagg atgaactcat tgaattcttt 600
tcccgagagg ctgacaatgt taaagacaaa ctttgcagta agcgaacaga tctttgtgac 660
catgccctgc acatatcgca tgatgagcta tgaaccactg gagcagccca cactggcttg 720
atggatcacc cccaggaggg gaaaatggtg gcaatgcctt ttatatatta tgtttttact 780
gaaattaact gaaaaaatat gaaaccaaaa gtac 814

<210> 6
<211> 182
<212> PRT
<213> Homo sapiens

<400> 6
Met Lys Gly Trp Gly Trp Leu Ala Leu Leu Leu Gly Ala Leu Leu Gly
1 5 10 15

Thr Ala Trp Ala Arg Arg Ser Gln Asp Leu His Cys Gly Ala Cys Arg
20 25 30

Ala Leu Val Asp Glu Leu Glu Trp Glu Ile Ala Gln Val Asp Pro Lys
35 40 45

Lys Thr Ile Gln Met Gly Ser Phe Arg Ile Asn Pro Asp Gly Ser Gln
50 55 60

Ser Val Val Glu Val Pro Tyr Ala Arg Ser Glu Ala His Leu Thr Glu
65 70 75 80

Leu Leu Glu Glu Ile Cys Asp Arg Met Lys Glu Tyr Gly Glu Gln Ile
85 90 95

Asp Pro Ser Thr His Arg Lys Asn Tyr Val Arg Val Val Gly Arg Asn
100 105 110

Gly Glu Ser Ser Glu Leu Asp Leu Gln Gly Ile Arg Ile Asp Ser Asp

115 120 125

Ile Ser Gly Thr Leu Lys Phe Ala Cys Glu Ser Ile Val Glu Glu Tyr
130 135 140

Glu Asp Glu Leu Ile Glu Phe Phe Ser Arg Glu Ala Asp Asn Val Lys
145 150 155 160

Asp Lys Leu Cys Ser Lys Arg Thr Asp Leu Cys Asp His Ala Leu His
165 170 175

Ile Ser His Asp Glu Leu
180

<210> 7
<211> 1597
<212> DNA
<213> Homo sapiens

<400> 7
gcggccgctg cgccgcaaac tcgtgtggga cgcaccgctc cagccgcccg cgggccagcg 60
caccggtccc ccagcggcag ccgagcccgc ccgcgcgccg ttcgtgccct cgtgaggctg 120
gcatgcagga tggcaggaca gcccggccac atgccccatg gagggagttc caacaacctc 180
tgccacaccc tggggcctgt gcatcctcct gacccacaga ggcatcccaa cacgctgtct 240
tttcgctgct cgctggcgga cttccagatc gaaaagaaga taggccgagg acagttcagc 300
gaggtgtaca aggccacctg cctgctggac aggaagacag tggctctgaa gaaggtgcag 360
atctttgaga tgatggacgc caaggcgagg caggactgtg tcaaggagat cggcctcttg 420
aagcaactga accacccaaa tatcatcaag tatttggact cgtttatcga agacaacgag 480
ctgaacattg tgctggagtt ggctgacgca ggggacctct cgcagatgat caagtacttt 540
aagaagcaga agcggctcat cccggagagg acagtatgga agtactttgt gcagctgtgc 600
agcgccgtgg agcacatgca ttcacgccgg gtgatgcacc gagacatcaa gcctgccaac 660
gtgttcatca cagccacggg cgtcgtgaag ctcggtgacc ttggtctggg ccgcttcttc 720
agctctgaga ccaccgcagc ccactcccta gtggggacgc cctactacat gtcaccggag 780
aggatccatg agaacggcta caacttcaag tccgacatct ggtccttggg ctgtctgctg 840
tacgagatgg cagccctcca gagccccttc tatggagata agatgaatct cttctccctg 900
tgccagaaga tcgagcagtg tgactacccc ccactccccg gggagcacta ctccgagaag 960
ttacgagaac tggtcagcat gtgcatctgc cctgaccccc accagagacc tgacatcgga 1020
tacgtgcacc aggtggccaa gcagatgcac atctggatgt ccagcacctg agcgtggatg 1080
caccgtgcct tatcaaagcc agcaccactt tgccttactt gagtcgtctt ctcttcgagt 1140
ggccacctgg tagcctagaa cagctaagac cacagggttc agcaggttcc ccaaaaggct 1200
gcccagcctt acagcagatg ctgaaggcag agcagctgag ggaggggcgc tggccacatg 1260
tcactgatgg tcagattcca aagtcctttc tttatactgt tgtggacaat ctcagctggg 1320
tcaataaggg caggtggttc agcgagccac ggcagccccc tgtatctgga ttgtaatgtg 1380
aatctttagg gtaattcctc cagtgacctg tcaaggctta tgctaacagg agacttgcag 1440
gagaccgtgt gatttgtgta gtgagccttt gaaaatggtt agtaccgggt tcagtttagt 1500
tcttggtatc ttttcaatca agctgtgtgc ttaatttact ctgttgtaaa gggataaagt 1560
ggaaatcatt tttttccgtg gaaaaaaaaa aaaaaaa 1597

<210> 8
<211> 306
<212> PRT
<213> Homo sapiens

400> 8
Met Pro His Gly Gly Ser Ser Asn Asn Leu Cys His Thr Leu Gly Pro
1 5 10 15

Val His Pro Pro Asp Pro Gln Arg His Pro Asn Thr Leu Ser Phe Arg

20 25 30

Cys Ser Leu Ala Asp Phe Gln Ile Glu Lys Lys Ile Gly Arg Gly Gln
35 40 45

Phe Ser Glu Val Tyr Lys Ala Thr Cys Leu Leu Asp Arg Lys Thr Val
50 55 60

Ala Leu Lys Lys Val Gln Ile Phe Glu Met Met Asp Ala Lys Ala Arg
65 70 75 80

Gln Asp Cys Val Lys Glu Ile Gly Leu Leu Lys Gln Leu Asn His Pro
85 90 95

Asn Ile Ile Lys Tyr Leu Asp Ser Phe Ile Glu Asp Asn Glu Leu Asn
100 105 110

Ile Val Leu Glu Leu Ala Asp Ala Gly Asp Leu Ser Gln Met Ile Lys
115 120 125

Tyr Phe Lys Lys Gln Lys Arg Leu Ile Pro Glu Arg Thr Val Trp Lys
130 135 140

Tyr Phe Val Gln Leu Cys Ser Ala Val Glu His Met His Ser Arg Arg
145 150 155 160

Val Met His Arg Asp Ile Lys Pro Ala Asn Val Phe Ile Thr Ala Thr
165 170 175

Gly Val Val Lys Leu Gly Asp Leu Gly Leu Gly Arg Phe Phe Ser Ser
180 185 190

Glu Thr Thr Ala Ala His Ser Leu Val Gly Thr Pro Tyr Tyr Met Ser
195 200 205

Pro Glu Arg Ile His Glu Asn Gly Tyr Asn Phe Lys Ser Asp Ile Trp
210 215 220

Ser Leu Gly Cys Leu Leu Tyr Glu Met Ala Ala Leu Gln Ser Pro Phe
225 230 235 240

Tyr Gly Asp Lys Met Asn Leu Phe Ser Leu Cys Gln Lys Ile Glu Gln
245 250 255

Cys Asp Tyr Pro Pro Leu Pro Gly Glu His Tyr Ser Glu Lys Leu Arg
260 265 270

Glu Leu Val Ser Met Cys Ile Cys Pro Asp Pro His Gln Arg Pro Asp
275 280 285

Ile Gly Tyr Val His Gln Val Ala Lys Gln Met His Ile Trp Met Ser
290 295 300

Ser Thr
305

<210> 9
<211> 6780
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> n=a,t,g or c

<400> 9
gggcggggct gagggcggcg ggggcgggcc gcccgagctg ggagggcggc ggcgccgagg 60
ggaggagagc ggcccatgga cccgcggggc ccggcgcccc agactctgcg ccgtcgggac 120
ggagcccaag atgtcggcct aggccggggc gcgacgacgc ggacggggcg gcgaggaggc 180
gccgctgctg ccggggctcg cagccgccga gcccccgagg gcgcgccctg acggactggc 240
cgagccggcg gtgagaggcc ggcgcgtctg gagcgggccg cgcggcacca tgtcggccaa 300
ggtgcggctc aagaagctgg agcagctgct cctggacggg ccctggcgca acgagagcgc 360
cctgagcgtg gaaacgctgc tcgacgtgct cgtctgcctg tacaccgagt gcagccactc 420
ggccctgcgc cgcgacaagt acgtggccga gttcctcgag tgggctaaac catttacaca 480
gctggtgaaa gaaatgcagc ttcatcgaga agactttgaa ataattaaag taattggaag 540
aggtgctttt ggtgaggttg ctgttgtcaa aatgaagaat actgaacgaa tttatgcaat 600
gaaaatcctc aacaagtggg agatgctgaa aagagcagag accgcgtgct tccgagagga 660
gcgcgatgtg ctggtgaacg gcgactgcca gtggatcacc gcgctgcact acgcctttca 720
ggacgagaac cacctgtact tagtcatgga ttactatgtg ggtggtgatt tactgaccct 780
gctcagcaaa tttgaagaca agcttccgga agatatggcg aggttctaca ttggtgaaat 840
ggtgctggcc attgactcca tccatcagct tcattacgtg cacagagaca ttaaacctga 900
caatgtcctt ttggacgtga atggtcatat ccgcctggct gactttggat catgtttgaa 960
gatgaatgat gatggcactg tgcagtcctc cgtggccgtg ggcacacctg actacatctc 1020
gccggagatc ctgcaggcga tggaggacgg catgggcaaa tacgggcctg agtgtgactg 1080
gtggtctctg ggtgtctgca tgtatgagat gctctatgga gaaacgccgt tttatgcgga 1140
gtcactcgtg gagacctatg ggaagatcat gaaccatgaa gagcgattcc agttcccatc 1200
ccatgtcacg gatgtatctg aagaagcgaa ggacctcatc cagagactga tctgcagtag 1260
agaacgccgg ctggggcaga atggaataga ggatttcaaa aagcatgcgt tttttgaagg 1320
tctaaattgg gaaaatatac gaaacctaga agcaccttat attcctgatg tgagcagtcc 1380
ctctgacaca tccaacttcg acgtggatga cgacgtgctg agaaacacgg aaatattacc 1440
tcctggttct cacacaggct tttctggatt acatttgcca ttcattggtt ttacattcac 1500
aacggaaagc tgtttttctg atcgaggctc tctgaagagc ataatgcagt ccaacacatt 1560
aaccaaagat gaggatgtgc agcgggacct ggagcacagc ctgcagatgg aagcttacga 1620
gaggaggatt cggaggctgg aacaggagaa gctggagctg agcaggaagc tgcaagagtc 1680
cacccagacc gtgcagtccc tccacggctc atctcgggcc ctcagcaatt caaaccgaga 1740
taaagaaatc aaaaagctaa atgaagaaat cgaacgcttg aagaataaaa tagcagattc 1800
aaacaggctg gagcgacagc ttgaggacac agtggcgctt cgccaagagc gtgaggactc 1860
cacgcagcgg ctgcgggggc tggagaagca gcaccgcgtg gtccggcagg agaaggagga 1920
gctgcacaag caactggttg aagcctcaga gcggttgaaa tcccaggcca aggaactcaa 1980
agatgcccat cagcagcgaa agctggccct gcaggagttc tcggagctga acgagcgcat 2040
ggcagagctc cgtgcccaga agcagaaggt gtcccggcag ctgcgagaca aggaggagga 2100
gatggaggtg gccacgcaga aggtggacgc catgcggcag gaaatgcgga gagctgagaa 2160
gctcaggaaa gagctggaag ctcagcttga tgatgctgtt gctgaggcct ccaaggagcg 2220
caagcttcgt gagcacagcg agaacttctg caagcaaatg gaaagcgagc tggaggccct 2280
caaggtgaag caaggaggcc ggggagcggg tgccacctta gagcaccagc aagagatttc 2340
caaaatcaaa tccgagctgg agaagaaagt cttattttat gaagaggaat tggtcagacg 2400
tgaggcctcc catgtgctag aagtgaaaaa tgtgaagaag gaggtgcatg attcagaaag 2460
ccaccagctg gccctgcaga aagaaatctt gatgttaaaa gataagttag aaaagtcaaa 2520
gcgagaacgg cataacgaga tggaggaggc agtaggtaca ataaaagata aatacgaacg 2580
agaaagagcg atgctgtttg atgaaaacaa gaagctaact gctgaaaatg aaaagctctg 2640
ttcctttgtg gataaactca cagctcaaaa tagacagctg gaggatgagc tgcaggatct 2700
ggcagccaag aaggagtcag tggcccactg ggaagctcag attgcggaaa tcattcagtg 2760
ggtcagtgac gagaaagatg cccggggtta ccttcaagct cttgcttcca agatgaccga 2820
agagctcgag gctttgagga gttctagtct ggggtcaaga acactggacc cgctgtggaa 2880
ggtgcgccgc agccagaagc tggacatgtc cgcgcggctg gagctgcagt cggccctgga 2940
ggcggagatc cgggccaagc agcttgtcca ggaggagctc aggaaggtca aggacgccaa 3000
cctcaccttg gaaagcaaac taaaggattc cgaagccaaa aacagagaat tattagaaga 3060
aatggaaatt ttgaagaaaa agatggaaga aaaattcaga gcagatactg ggctcaaact 3120

tccagatttt caggattcca tttttgagta tttcaacact gctcctcttg cacatgacct 3180
gacatttaga accagctcag ctagtgagca agaaacacaa gctccgaagc cagaagcgtc 3240
cccgtcgatg tctgtggctg catcagagca gcaggaggac atggctcggc ccccgcagag 3300
gccatccgct gtgccgttgc ccaccacgca ggccctggtt ctggctggac cgaagccaaa 3360
agctcaccag ttcagcatca agtccttctc cagccctact cagtgcagcc actgcacctc 3420
cctgatggtt gggctgatcc ggcagggcta cgcctgcgag gtgtgttcct ttgcttgcca 3480
cgtgtcctgc aaagacggtg ccccccaggt gtgcccaata cctcccgagc agtccaagag 3540
gcctctgggc gtggacgtgc agcgaggcat cggaacagcc tacaaaggcc atgtcaaggt 3600
cccaaagccc acgggggtga agaagggatg gcagcgcgca tatgcagtcg tctgtgagtg 3660
caagctcttc ctgtatgatc tgcctgaagg aaaatccacc cagcctggtg tcattgcgag 3720
ccaagtcttg gatctcagag atgacgagtt ttccgtgagc tcagtcctgg cctcagatgt 3780
cattcatgct acacgccgag atattccatg tatattcagg gtgacggcct ctctcttagg 3840
tgcaccttct aagaccagct cgctgctcat tctgacagaa aatgagaatg aaaagaggaa 3900
gtgggttggg attctagaag gactccagtc catccttcat aaaaaccggc tgaggaatca 3960
ggtcgtgcat gttcccttgg aagcctacga cagctcgctg cctctcatca aggccatcct 4020
gacagctgcc atcgtggatg cagacaggat tgcagtcggc ctagaagaag ggctctatgt 4080
catagaggtc acccgagatg tgatcgtccg tgccgctgac tgtaagaagg tacaccagat 4140
cgagcttgct cccagggaga agatcgtaat cctcctctgt ggccggaacc accatgtgca 4200
cctctatccg tggtcgtccc ttgatggagc ggaaggcagc tttgacatca agcttccgga 4260
aaccaaaggc tgccagctca tggccacggc cacactcaag aggaactctg gcacctgcct 4320
gtttgtggcc gtgaaacggc tgatcctttg ctatgagatc cagagaacga agccattcca 4380
cagaaagttc aatgagattg tggctcccgg cagcgtgcag tgcctggcgg tgctcaggga 4440
caggctctgt gtgggctacc cttctgggtt ctgcctgctg agcatccagg gggacgggca 4500
gcctctaaac ctggtaaatc ccaatgaccc ctcgcttgcg ttcctctcac aacagtcttt 4560
tgatgccctt tgtgctgtgg agctcgaaag cgaggagtac ctgctttgct tcagccacat 4620
gggactgtac gtggacccgc aaggccggag ggcacgcgcg caggagctca tgtggcctgc 4680
ggctcctgtc gcctgtagtt gcagccccac ccacgtcacg gtgtacagcg agtatggcgt 4740
ggacgtcttt gatgtgcgca ccatggagtg ggtgcagacc atcggcctgc ggaggataag 4800
gcccctgaac tctgaaggca ccctcaacct cctcaactgc gagcctccac gcttgatcta 4860
cttcaagagc aagttctcgg gagcggttct caacgtgccg gacacctccg acaacagcaa 4920
gaagcagatg ctgcgcacca ggagcaaaag gcggttcgtc ttcaaggtcc cagaggaaga 4980
gagactgcag cagaggcgag agatgcttag agacccagaa ttgagatcca aaatgatatc 5040
caacccaacc aacttcaacc acgtggccca catgggccca ggcgacggca tgcaggtgct 5100
catggacctg cctctgagtg ctgtgccccc ctcccaggag gaaaggccgg gccccgctcc 5160
caccaacctg gctcgccagc ctccatccag gaacaagccc tacatctcgt ggccctcatc 5220
aggtggatcg gagcctagcg tgactgtgcc tctgagaagt atgtctgatc cagaccagga 5280
ctttgacaaa gagcctgatt cggactccac caaacactca actccatcga atagctccaa 5340
ccccagcggc ccaccgagcc ccaactcccc ccacaggagc cagctccccc tcgaaggcct 5400
ggagcagccg gcctgtgaca cctgaagccg ccagctcgcc acaggggcca gggagctgga 5460
gatggcctcc agcgtcagtg ccaagactga gcgggccctc cagtgttgtc caaggaaatg 5520
tagaatcact ttgtagatat ggagatgaag aagacaaatc tttattataa tattgatcag 5580
ttttatgccg cattgttcgt ggcagtagac cacatctgtt cgtctgcaca gctgtgaggc 5640
gatgctgttc catctgcaca tgaaggaccc ccatacagcc tgtctcccac ccctgacaac 5700
ccgagagggc atatggggcc ctgccaacac cacttcctca gcagaaaccc gtcatgacgc 5760
ggctgcttcg gaagcagaca tctggggaca cagcctcagt acccagtctt ttccctagtt 5820
cctgaaactt tcctaggacc ttaagagaat agtaggaggt cctatagcat tcccagtgtc 5880
actagaattt tgaagacagg aaagtggagg ttagtctgtg gccttttttt catttagcca 5940
ttgcacagtc agctgcagaa gtcctgctga ccacctagtc atggacaaag gcccaggacc 6000
agtgacaccc tgcgtccctg tgtgcgttaa gttcattctg ggtcgcagcc atgaagtgtc 6060
accagtatct actactgtga agtcagctgt gctgttttcc attcgcttcc acggcttctg 6120
cctcctgcca taaaaccagc gagtgtcgtg gtgcaggcag gccctgtggc ctgctgggct 6180
gagggaagtc agagccccag ggcgccacga agcagccact gggatacccc accccgcccc 6240
gccnnccccc ccccccccnc cagtcnagnc ccgaaatgga gcccccgtga ttagtagccc 6300
gtatgatcac gtagacccac ccaacacact cctgcacact ggccccggcc cacggcacag 6360
caatcccctg cgcgtggatt tcacctcacc ctttgtacca gatgttgagt gaccagctct 6420
gtggccctgt gtcgtcagag gcttgtgatt aactgtggcg gcagacacag cttgtccaca 6480
gcttgggcca ggcttccccc gtcctcccac cggtcggctg cttggcaagg ctgttcagga 6540
cgtgcacttc cccaagtcgg cactgagtgg cccagcacca cctagccctg ccaccccact 6600
gccctcctgg gccttctgct ggatgggcac ctgggggttt ctggtttttt actttttaa 6660

tgtaagtctc agtctttgta attaattatt gaattgtgag aacatttttg aacaatttac 6720
ctgtcaataa agcagaagac ggcagtttta aagttaaaaa aaaaaaaaaa aaaaaaaaaa 6780

<210> 10
<211> 1711
<212> PRT
<213> Homo sapiens

<400> 10
Met Ser Ala Lys Val Arg Leu Lys Lys Leu Glu Gln Leu Leu Leu Asp
1 5 10 15

Gly Pro Trp Arg Asn Glu Ser Ala Leu Ser Val Glu Thr Leu Leu Asp
20 25 30

Val Leu Val Cys Leu Tyr Thr Glu Cys Ser His Ser Ala Leu Arg Arg
35 40 45

Asp Lys Tyr Val Ala Glu Phe Leu Glu Trp Ala Lys Pro Phe Thr Gln
50 55 60

Leu Val Lys Glu Met Gln Leu His Arg Glu Asp Phe Glu Ile Ile Lys
65 70 75 80

Val Ile Gly Arg Gly Ala Phe Gly Glu Val Ala Val Val Lys Met Lys
85 90 95

Asn Thr Glu Arg Ile Tyr Ala Met Lys Ile Leu Asn Lys Trp Glu Met
100 105 110

Leu Lys Arg Ala Glu Thr Ala Cys Phe Arg Glu Glu Arg Asp Val Leu
115 120 125

Val Asn Gly Asp Cys Gln Trp Ile Thr Ala Leu His Tyr Ala Phe Gln
130 135 140

Asp Glu Asn His Leu Tyr Leu Val Met Asp Tyr Tyr Val Gly Gly Asp
145 150 155 160

Leu Leu Thr Leu Leu Ser Lys Phe Glu Asp Lys Leu Pro Glu Asp Met
165 170 175

Ala Arg Phe Tyr Ile Gly Glu Met Val Leu Ala Ile Asp Ser Ile His
180 185 190

Gln Leu His Tyr Val His Arg Asp Ile Lys Pro Asp Asn Val Leu Leu
195 200 205

Asp Val Asn Gly His Ile Arg Leu Ala Asp Phe Gly Ser Cys Leu Lys
210 215 220

Met Asn Asp Asp Gly Thr Val Gln Ser Ser Val Ala Val Gly Thr Pro
225 230 235 240

Asp Tyr Ile Ser Pro Glu Ile Leu Gln Ala Met Glu Asp Gly Met Gly
245 250 255

Lys Tyr Gly Pro Glu Cys Asp Trp Trp Ser Leu Gly Val Cys Met Tyr

260 265 270

Glu Met Leu Tyr Gly Glu Thr Pro Phe Tyr Ala Glu Ser Leu Val Glu
275 280 285

Thr Tyr Gly Lys Ile Met Asn His Glu Glu Arg Phe Gln Phe Pro Ser
290 295 300

His Val Thr Asp Val Ser Glu Glu Ala Lys Asp Leu Ile Gln Arg Leu
305 310 315 320

Ile Cys Ser Arg Glu Arg Arg Leu Gly Gln Asn Gly Ile Glu Asp Phe
325 330 335

Lys Lys His Ala Phe Phe Glu Gly Leu Asn Trp Glu Asn Ile Arg Asn
340 345 350

Leu Glu Ala Pro Tyr Ile Pro Asp Val Ser Ser Pro Ser Asp Thr Ser
355 360 365

Asn Phe Asp Val Asp Asp Asp Val Leu Arg Asn Thr Glu Ile Leu Pro
370 375 380

Pro Gly Ser His Thr Gly Phe Ser Gly Leu His Leu Pro Phe Ile Gly
385 390 395 400

Phe Thr Phe Thr Thr Glu Ser Cys Phe Ser Asp Arg Gly Ser Leu Lys
405 410 415

Ser Ile Met Gln Ser Asn Thr Leu Thr Lys Asp Glu Asp Val Gln Arg
420 425 430

Asp Leu Glu His Ser Leu Gln Met Glu Ala Tyr Glu Arg Arg Ile Arg
435 440 445

Arg Leu Glu Gln Glu Lys Leu Glu Leu Ser Arg Lys Leu Gln Glu Ser
450 455 460

Thr Gln Thr Val Gln Ser Leu His Gly Ser Ser Arg Ala Leu Ser Asn
465 470 475 480

Ser Asn Arg Asp Lys Glu Ile Lys Lys Leu Asn Glu Glu Ile Glu Arg
485 490 495

Leu Lys Asn Lys Ile Ala Asp Ser Asn Arg Leu Glu Arg Gln Leu Glu
500 505 510

Asp Thr Val Ala Leu Arg Gln Glu Arg Glu Asp Ser Thr Gln Arg Leu
515 520 525

Arg Gly Leu Glu Lys Gln His Arg Val Val Arg Gln Glu Lys Glu Glu
530 535 540

Leu His Lys Gln Leu Val Glu Ala Ser Glu Arg Leu Lys Ser Gln Ala
545 550 555 560

Lys Glu Leu Lys Asp Ala His Gln Gln Arg Lys Leu Ala Leu Gln Glu
565 570 575

Phe Ser Glu Leu Asn Glu Arg Met Ala Glu Leu Arg Ala Gln Lys Gln
580 585 590

Lys Val Ser Arg Gln Leu Arg Asp Lys Glu Glu Glu Met Glu Val Ala
595 600 605

Thr Gln Lys Val Asp Ala Met Arg Gln Glu Met Arg Arg Ala Glu Lys
610 615 620

Leu Arg Lys Glu Leu Glu Ala Gln Leu Asp Asp Ala Val Ala Glu Ala
625 630 635 640

Ser Lys Glu Arg Lys Leu Arg Glu His Ser Glu Asn Phe Cys Lys Gln
645 650 655

Met Glu Ser Glu Leu Glu Ala Leu Lys Val Lys Gln Gly Gly Arg Gly
660 665 670

Ala Gly Ala Thr Leu Glu His Gln Gln Glu Ile Ser Lys Ile Lys Ser
675 680 685

Glu Leu Glu Lys Lys Val Leu Phe Tyr Glu Glu Glu Leu Val Arg Arg
690 695 700

Glu Ala Ser His Val Leu Glu Val Lys Asn Val Lys Lys Glu Val His
705 710 715 720

Asp Ser Glu Ser His Gln Leu Ala Leu Gln Lys Glu Ile Leu Met Leu
725 730 735

Lys Asp Lys Leu Glu Lys Ser Lys Arg Glu Arg His Asn Glu Met Glu
740 745 750

Glu Ala Val Gly Thr Ile Lys Asp Lys Tyr Glu Arg Glu Arg Ala Met
755 760 765

Leu Phe Asp Glu Asn Lys Lys Leu Thr Ala Glu Asn Glu Lys Leu Cys
770 775 780

Ser Phe Val Asp Lys Leu Thr Ala Gln Asn Arg Gln Leu Glu Asp Glu
785 790 795 800

Leu Gln Asp Leu Ala Ala Lys Lys Glu Ser Val Ala His Trp Glu Ala
805 810 815

Gln Ile Ala Glu Ile Ile Gln Trp Val Ser Asp Glu Lys Asp Ala Arg
820 825 830

Gly Tyr Leu Gln Ala Leu Ala Ser Lys Met Thr Glu Glu Leu Glu Ala
835 840 845

Leu Arg Ser Ser Ser Leu Gly Ser Arg Thr Leu Asp Pro Leu Trp Lys
850 855 860

Val Arg Arg Ser Gln Lys Leu Asp Met Ser Ala Arg Leu Glu Leu Gln
865 870 875 880

Ser Ala Leu Glu Ala Glu Ile Arg Ala Lys Gln Leu Val Gln Glu Glu
885 890 895

Leu Arg Lys Val Lys Asp Ala Asn Leu Thr Leu Glu Ser Lys Leu Lys
900 905 910

Asp Ser Glu Ala Lys Asn Arg Glu Leu Leu Glu Glu Met Glu Ile Leu
915 920 925

Lys Lys Lys Met Glu Glu Lys Phe Arg Ala Asp Thr Gly Leu Lys Leu
930 935 940

Pro Asp Phe Gln Asp Ser Ile Phe Glu Tyr Phe Asn Thr Ala Pro Leu
945 950 955 960

Ala His Asp Leu Thr Phe Arg Thr Ser Ser Ala Ser Glu Gln Glu Thr
965 970 975

Gln Ala Pro Lys Pro Glu Ala Ser Pro Ser Met Ser Val Ala Ala Ser
980 985 990

Glu Gln Gln Glu Asp Met Ala Arg Pro Pro Gln Arg Pro Ser Ala Val
995 1000 1005

Pro Leu Pro Thr Thr Gln Ala Leu Val Leu Ala Gly Pro Lys Pro
1010 1015 1020

Lys Ala His Gln Phe Ser Ile Lys Ser Phe Ser Ser Pro Thr Gln
1025 1030 1035

Cys Ser His Cys Thr Ser Leu Met Val Gly Leu Ile Arg Gln Gly
1040 1045 1050

Tyr Ala Cys Glu Val Cys Ser Phe Ala Cys His Val Ser Cys Lys
1055 1060 1065

Asp Gly Ala Pro Gln Val Cys Pro Ile Pro Pro Glu Gln Ser Lys
1070 1075 1080

Arg Pro Leu Gly Val Asp Val Gln Arg Gly Ile Gly Thr Ala Tyr
1085 1090 1095

Lys Gly His Val Lys Val Pro Lys Pro Thr Gly Val Lys Lys Gly
1100 1105 1110

Trp Gln Arg Ala Tyr Ala Val Val Cys Glu Cys Lys Leu Phe Leu
1115 1120 1125

Tyr Asp Leu Pro Glu Gly Lys Ser Thr Gln Pro Gly Val Ile Ala
1130 1135 1140

Ser Gln Val Leu Asp Leu Arg Asp Asp Glu Phe Ser Val Ser Ser
1145 1150 1155

Val Leu Ala Ser Asp Val Ile His Ala Thr Arg Arg Asp Ile Pro
1160 1165 1170

Cys Ile Phe Arg Val Thr Ala Ser Leu Leu Gly Ala Pro Ser Lys
1175 1180 1185

Thr Ser Ser Leu Leu Ile Leu Thr Glu Asn Glu Asn Glu Lys Arg

1190 1195 1200

Lys Trp Val Gly Ile Leu Glu Gly Leu Gln Ser Ile Leu His Lys
1205 1210 1215

Asn Arg Leu Arg Asn Gln Val Val His Val Pro Leu Glu Ala Tyr
1220 1225 1230

Asp Ser Ser Leu Pro Leu Ile Lys Ala Ile Leu Thr Ala Ala Ile
1235 1240 1245

Val Asp Ala Asp Arg Ile Ala Val Gly Leu Glu Glu Gly Leu Tyr
1250 1255 1260

Val Ile Glu Val Thr Arg Asp Val Ile Val Arg Ala Ala Asp Cys
1265 1270 1275

Lys Lys Val His Gln Ile Glu Leu Ala Pro Arg Glu Lys Ile Val
1280 1285 1290

Ile Leu Leu Cys Gly Arg Asn His His Val His Leu Tyr Pro Trp
1295 1300 1305

Ser Ser Leu Asp Gly Ala Glu Gly Ser Phe Asp Ile Lys Leu Pro
1310 1315 1320

Glu Thr Lys Gly Cys Gln Leu Met Ala Thr Ala Thr Leu Lys Arg
1325 1330 1335

Asn Ser Gly Thr Cys Leu Phe Val Ala Val Lys Arg Leu Ile Leu
1340 1345 1350

Cys Tyr Glu Ile Gln Arg Thr Lys Pro Phe His Arg Lys Phe Asn
1355 1360 1365

Glu Ile Val Ala Pro Gly Ser Val Gln Cys Leu Ala Val Leu Arg
1370 1375 1380

Asp Arg Leu Cys Val Gly Tyr Pro Ser Gly Phe Cys Leu Leu Ser
1385 1390 1395

Ile Gln Gly Asp Gly Gln Pro Leu Asn Leu Val Asn Pro Asn Asp
1400 1405 1410

Pro Ser Leu Ala Phe Leu Ser Gln Gln Ser Phe Asp Ala Leu Cys
1415 1420 1425

Ala Val Glu Leu Glu Ser Glu Glu Tyr Leu Leu Cys Phe Ser His
1430 1435 1440

Met Gly Leu Tyr Val Asp Pro Gln Gly Arg Arg Ala Arg Ala Gln
1445 1450 1455

Glu Leu Met Trp Pro Ala Ala Pro Val Ala Cys Ser Cys Ser Pro
1460 1465 1470

Thr His Val Thr Val Tyr Ser Glu Tyr Gly Val Asp Val Phe Asp
1475 1480 1485

Val Arg Thr Met Glu Trp Val Gln Thr Ile Gly Leu Arg Arg Ile
1490 1495 1500

Arg Pro Leu Asn Ser Glu Gly Thr Leu Asn Leu Leu Asn Cys Glu
1505 1510 1515

Pro Pro Arg Leu Ile Tyr Phe Lys Ser Lys Phe Ser Gly Ala Val
1520 1525 1530

Leu Asn Val Pro Asp Thr Ser Asp Asn Ser Lys Lys Gln Met Leu
1535 1540 1545

Arg Thr Arg Ser Lys Arg Arg Phe Val Phe Lys Val Pro Glu Glu
1550 1555 1560

Glu Arg Leu Gln Gln Arg Arg Glu Met Leu Arg Asp Pro Glu Leu
1565 1570 1575

Arg Ser Lys Met Ile Ser Asn Pro Thr Asn Phe Asn His Val Ala
1580 1585 1590

His Met Gly Pro Gly Asp Gly Met Gln Val Leu Met Asp Leu Pro
1595 1600 1605

Leu Ser Ala Val Pro Pro Ser Gln Glu Glu Arg Pro Gly Pro Ala
1610 1615 1620

Pro Thr Asn Leu Ala Arg Gln Pro Pro Ser Arg Asn Lys Pro Tyr
1625 1630 1635

Ile Ser Trp Pro Ser Ser Gly Gly Ser Glu Pro Ser Val Thr Val
1640 1645 1650

Pro Leu Arg Ser Met Ser Asp Pro Asp Gln Asp Phe Asp Lys Glu
1655 1660 1665

Pro Asp Ser Asp Ser Thr Lys His Ser Thr Pro Ser Asn Ser Ser
1670 1675 1680

Asn Pro Ser Gly Pro Pro Ser Pro Asn Ser Pro His Arg Ser Gln
1685 1690 1695

Leu Pro Leu Glu Gly Leu Glu Gln Pro Ala Cys Asp Thr
1700 1705 1710

<210> 11
<211> 2287
<212> DNA
<213> Homo sapiens

<400> 11
ggctgaggca gtggctcctt gcacagcagc tgcacgcgcc gtggctccgg atcttcttcg 60
tctttgcagc gtagcccgag tcggtcagcg ccagaggacc tcagcagcca tgtcgaagcc 120
ccatagtgaa gccgggactg ccttcattca gacccagcag ctgcacgcag ccatggctga 180
cacattcctg gagcacatgt gccgcctgga cattgattca ccacccatca cagcccggaa 240
cactggcatc atctgtacca ttggcccagc ttcccgatca gtggagacgt tgaaggagat 300
gattaagtct ggaatgaatg tggctcgtct gaacttctct catggaactc atgagtacca 360
tgcggagacc atcaagaatg tgcgcacagc cacggaaagc tttgcttctg accccatcct 420

ctaccggccc gttgctgtgg ctctagacac taaaggacct gagatccgaa ctgggctcat 480
caagggcagc ggcactgcag aggtggagct gaagaaggga gccactctca aaatcacgct 540
ggataacgcc tacatggaaa agtgtgacga gaacatcctg tggctggact acaagaacat 600
ctgcaaggtg gtggaagtgg gcagcaagat ctacgtggat gatgggctta tttctctcca 660
ggtgaagcag aaaggtgccg acttcctggt gacggaggtg gaaaatggtg gctccttggg 720
cagcaagaag ggtgtgaacc ttcctggggc tgctgtggac ttgcctgctg tgtcggagaa 780
ggacatccag gatctgaagt ttggggtcga gcaggatgtt gatatggtgt ttgcgtcatt 840
catccgcaag gcatctgatg tccatgaagt taggaaggtc ctgggagaga agggaaagaa 900
catcaagatt atcagcaaaa tcgagaatca tgaggggttt cggaggtttg atgaaatcct 960
ggaggccagt gatgggatca tggtggctcg tggtgatcta ggcattgaga ttcctgcaga 1020
gaaggtcttc cttgctcaga agatgatgat tggacggtgc aaccgagctg ggaagcctgt 1080
catctgtgct actcagatgc tggagagcat gatcaagaag ccccgcccca ctcgggctga 1140
aggcagtgat gtggccaatg cagtcctgga tggagccgac tgcatcatgc tgtctggaga 1200
aacagccaaa ggggactatc ctctggaggc tgtgcgcatg cagaacctga ttgcccgtga 1260
ggcagaggct gccatctacc acttgcaatt atttgaggaa ctccgccgcc tggcgcccat 1320
taccagcgac cccacagaag ccaccgccgt gggtgccgtg gaggcctcct tcaagtgctg 1380
cagtggggcc ataatcgtcc tcaccaagtc tggcaggtct gctcaccagg tggccagata 1440
ccgcccacgt gcccccatca ttgctgtgac ccggaatccc cagacagctc gtcaggccca 1500
cctgtaccgt ggcatcttcc ctgtgctgtg caaggaccca gtccaggagg cctgggctga 1560
ggacgtggac ctccgggtga actttgccat gaatgttggc aaggcccgag gcttcttcaa 1620
gaagggagat gtggtcattg tgctgaccgg atggcgccct ggctccggct tcaccaacac 1680
catgcgtgtt gttcctgtgc cgtgatggac cccagagccc ctcctccagc ccctgtccca 1740
ccccttccc ccagcccatc cattaggcca gcaacgcttg tagaactcac tctgggctgt 1800
aacgtggcac tggtaggttg ggacaccagg gaagaagatc aacgcctcac tgaaacatgg 1860
ctgtgtttgc agcctgctct agtgggacag cccagagcct ggctgcccca tcatgtggcc 1920
ccacccaatc aagggaagaa ggaggaatgc tggactggag gcccctggag ccagatggca 1980
agagggtgac agcttccttt cctgtgtgta ctctgtccag ttcctttaga aaaaatggat 2040
gcccagagga ctcccaaccc tggcttgggg tcaagaaaca gccagcaaga gttaggggcc 2100
ttagggcact gggctgttgt tccattgaag ccgactctgg ccctggccct tacttgcttc 2160
tctagctctc taggcctctc cagtttgcac ctgtccccac cctccactca gctgtcctgc 2220
agcaaacact ccacctcca ccttccattt tcccccacta ctgcagcacc tccaggcctg 2280
ttgccgc 2287

<210> 12
<211> 531
<212> PRT
<213> Homo sapiens

<400> 12
Met Ser Lys Pro His Ser Glu Ala Gly Thr Ala Phe Ile Gln Thr Gln
1 5 10 15

Gln Leu His Ala Ala Met Ala Asp Thr Phe Leu Glu His Met Cys Arg
20 25 30

Leu Asp Ile Asp Ser Pro Pro Ile Thr Ala Arg Asn Thr Gly Ile Ile
35 40 45

Cys Thr Ile Gly Pro Ala Ser Arg Ser Val Glu Thr Leu Lys Glu Met
50 55 60

Ile Lys Ser Gly Met Asn Val Ala Arg Leu Asn Phe Ser His Gly Thr
65 70 75 80

His Glu Tyr His Ala Glu Thr Ile Lys Asn Val Arg Thr Ala Thr Glu
85 90 95

Ser Phe Ala Ser Asp Pro Ile Leu Tyr Arg Pro Val Ala Val Ala Leu

```
          100                 105                 110

Asp Thr Lys Gly Pro Glu Ile Arg Thr Gly Leu Ile Lys Gly Ser Gly
        115                 120                 125

Thr Ala Glu Val Glu Leu Lys Lys Gly Ala Thr Leu Lys Ile Thr Leu
    130                 135                 140

Asp Asn Ala Tyr Met Glu Lys Cys Asp Glu Asn Ile Leu Trp Leu Asp
145                 150                 155                 160

Tyr Lys Asn Ile Cys Lys Val Val Glu Val Gly Ser Lys Ile Tyr Val
            165                 170                 175

Asp Asp Gly Leu Ile Ser Leu Gln Val Lys Gln Lys Gly Ala Asp Phe
            180                 185                 190

Leu Val Thr Glu Val Glu Asn Gly Gly Ser Leu Gly Ser Lys Lys Gly
        195                 200                 205

Val Asn Leu Pro Gly Ala Ala Val Asp Leu Pro Ala Val Ser Glu Lys
    210                 215                 220

Asp Ile Gln Asp Leu Lys Phe Gly Val Glu Gln Asp Val Asp Met Val
225                 230                 235                 240

Phe Ala Ser Phe Ile Arg Lys Ala Ser Asp Val His Glu Val Arg Lys
                245                 250                 255

Val Leu Gly Glu Lys Gly Lys Asn Ile Lys Ile Ile Ser Lys Ile Glu
            260                 265                 270

Asn His Glu Gly Val Arg Arg Phe Asp Glu Ile Leu Glu Ala Ser Asp
        275                 280                 285

Gly Ile Met Val Ala Arg Gly Asp Leu Gly Ile Glu Ile Pro Ala Glu
    290                 295                 300

Lys Val Phe Leu Ala Gln Lys Met Met Ile Gly Arg Cys Asn Arg Ala
305                 310                 315                 320

Gly Lys Pro Val Ile Cys Ala Thr Gln Met Leu Glu Ser Met Ile Lys
                325                 330                 335

Lys Pro Arg Pro Thr Arg Ala Glu Gly Ser Asp Val Ala Asn Ala Val
            340                 345                 350

Leu Asp Gly Ala Asp Cys Ile Met Leu Ser Gly Glu Thr Ala Lys Gly
        355                 360                 365

Asp Tyr Pro Leu Glu Ala Val Arg Met Gln Asn Leu Ile Ala Arg Glu
    370                 375                 380

Ala Glu Ala Ala Ile Tyr His Leu Gln Leu Phe Glu Glu Leu Arg Arg
385                 390                 395                 400

Leu Ala Pro Ile Thr Ser Asp Pro Thr Glu Ala Thr Ala Val Gly Ala
                405                 410                 415
```

Val Glu Ala Ser Phe Lys Cys Cys Ser Gly Ala Ile Ile Val Leu Thr
420 425 430

Lys Ser Gly Arg Ser Ala His Gln Val Ala Arg Tyr Arg Pro Arg Ala
435 440 445

Pro Ile Ile Ala Val Thr Arg Asn Pro Gln Thr Ala Arg Gln Ala His
450 455 460

Leu Tyr Arg Gly Ile Phe Pro Val Leu Cys Lys Asp Pro Val Gln Glu
465 470 475 480

Ala Trp Ala Glu Asp Val Asp Leu Arg Val Asn Phe Ala Met Asn Val
485 490 495

Gly Lys Ala Arg Gly Phe Phe Lys Lys Gly Asp Val Val Ile Val Leu
500 505 510

Thr Gly Trp Arg Pro Gly Ser Gly Phe Thr Asn Thr Met Arg Val Val
515 520 525

Pro Val Pro
530

<210> 13
<211> 4364
<212> DNA
<213> Homo sapiens

<400> 13
cattagatct ttacatgaaa gtaaaattta taagatttct agaaagtcaa aagatgataa 60
ctatttctta ggatactaaa agcactcaca ttatagaaaa aaaatcagtt aactatactc 120
cacaaacatt aaaggctccc tataaaaaaa catttttaat aggcaagcca cagaaagggc 180
aaatattaat agtttgcaat acatatgtat gaaaaggaat tgaatctaga atatttaaca 240
aagctttaca actcaaaaaa tacaaagaaa atatttttct tccaattggc aaattactta 300
aacagaacct tcacaaaaga agataagaat gtttaataaa catttgaagc cataataatg 360
acatcattag ccatgatgga aatgcaaatt taagtaccac ttcacatcca caagaaaaag 420
ataaaaataa aaggactgag ctcaccaaac attggtgagg atgtggtaat actgaaattc 480
ttgtaccgtg ctcctgaggg tataacatat tacaggattt ttttgaaaac tagtggttcc 540
ttataaactt aatgccctgg caacctcaca cctatttact taagaatgaa agggccccgc 600
cctcctccct cctcgctcgc gggccgggcc cggcatggtg cggcgtcgcc gccgatggcg 660
ctgaggcgga gcatggggcg gccggggctc ccgccgctgc cgctgccgcc gccaccgcgg 720
ctcgggctgc tgctggcgga gtccgccgcc gcaggtctga agctcatggg agccccggtg 780
aagctgacag tgtctcaggg gcagccggtg aagctcaact gcagtgtgga ggggatggag 840
gagcctgaca tccagtgggt gaaggatggg gctgtggtcc agaacttgga ccagttgtac 900
atcccagtca gcgagcagca ctggatcggc ttcctcagcc tgaagtcagt ggagcgctct 960
gacgccggcc ggtactggtg ccaggtggag gatgggggtg aaaccgagat ctcccagcca 1020
gtgtggctca cggtagaagg tgtgccattt ttcacagtgg agccaaaaga tctggcagtg 1080
ccacccaatg ccccttccca actgtcttgt gaggctgtgg gtccccctga acctgttacc 1140
attgtctggt ggagaggaac tacgaagatc gggggacccg ctccctctcc atctgtttta 1200
aatgtaacag gggtgaccca gagcaccatg ttttcctgtg aagctcacaa cctaaaaggc 1260
ctggcctctt ctcgcacagc cactgttcac cttcaagcac tgcctgcagc ccccttcaac 1320
atcaccgtga caaagctttc cagcagcaac gctagtgtgg cctggatgcc aggtgctgat 1380
ggccgagctc tgctacagtc ctgtacagtt caggtgacac aggccccagg aggctgggaa 1440
gtcctggctg ttgtggtccc tgtgcccccc tttacctgcc tgctccggga cctggtgcct 1500
gccaccaact acagcctcag ggtgcgctgt gccaatgcct tggggccctc tccctatgct 1560
gactgggtgc cctttcagac caagggtcta gccccagcca gcgctcccca aaacctccat 1620
gccatccgca cagattcagg cctcatcttg gagtgggaag aagtgatccc cgaggccccc 1680
ttggaaggcc ccctgggacc ctacaaactg tcctgggttc aagacaatgg aacccaggat 1740

gagctgacag tggaggggac cagggccaat ttgacaggct gggatcccca aaaggacctg 1800
atcgtacgtg tgtgcgtctc caatgcagtt ggctgtggac cctggagtca gccactggtg 1860
gtctcttctc atgaccgtgc aggccagcag ggccctcctc acagccgcac atcctgggta 1920
cctgtggtcc ttggtgtgct aacggccctg gtgacggctg ctgccctggc cctcatcctg 1980
cttcgaaaga gacggaaaga gacgcggttt gggcaagcct ttgacagtgt catggcccgg 2040
ggagagccag ccgttcactt ccgggcagcc cggtccttca atcgagaaag gcccgagcgc 2100
atcgaggcca cattggacag cttgggcatc agcgatgaac taaaggaaaa actggaggat 2160
gtgctcatcc cagagcagca gttcaccctg ggccggatgt tgggcaaagg agagtttggt 2220
tcagtgcggg aggcccagct gaagcaagag gatggctcct ttgtgaaagt ggctgtgaag 2280
atgctgaaag ctgacatcat tgcctcaagc gacattgaag agttcctcag ggaagcagct 2340
tgcatgaagg agtttgacca tccacacgtg gccaaacttg ttggggtaag cctccggagc 2400
agggctaaag gccgtctccc catccccatg gtcatcttgc ccttcatgaa gcatggggac 2460
ctgcatgcct tcctgctcgc ctcccggatt ggggagaacc cctttaacct acccctccag 2520
accctgatcc ggttcatggt ggacattgcc tgcggcatgg agtacctgag ctctcggaac 2580
ttcatccacc gagacctggc tgctcggaat tgcatgctgg cagaggacat gacagtgtgt 2640
gtggctgact tcggactctc ccggaagatc tacagtgggg actactatcg tcaaggctgt 2700
gcctccaaac tgcctgtcaa gtggctggcc ctggagagcc tggccgacaa cctgtatact 2760
gtgcagagtg acgtgtgggc gttcggggtg accatgtggg agatcatgac acgtgggcag 2820
acgccatatg ctggcatcga aaacgctgag atttacaact acctcattgg cgggaaccgc 2880
ctgaaacagc ctccggagtg tatggaggac gtgtatgatc tcatgtacca gtgctggagt 2940
gctgacccca agcagcgccc gagctttact tgtctgcgaa tggaactgga gaacatcttg 3000
ggccagctgt ctgtgctatc tgccagccag gaccccttat acatcaacat cgagagagct 3060
gaggagccca ctgtgggagg cagcctggag ctacctggca gggatcagcc ctacagtggg 3120
gctggggatg gcagtggcat gggggcagtg ggtggcactc ccagtgactg tcggtacata 3180
ctcacccccg gagggctggc tgagcagcca gggcaggcag agcaccagcc agagagtccc 3240
ctcaatgaga cacagaggct tttgctgctg cagcaagggc tactgccaca cagtagctgt 3300
tagcccacag gcagagggca tcggggccat ttggccggct ctggtggcca ctgagctggc 3360
tgactaagcc ccgtctgacc ccagcccaga cagcaaggtg tggaggctcc tgtggtagtc 3420
ctcccaagct gtgctgggaa gcccggactg accaaatcac ccaatcccag ttcttcctgc 3480
aaccactctg tggccagcct ggcatcagtt taggccttgg cttgatggaa gtgggccagt 3540
cctggttgtc tgaacccagg cagctggcag gagtggggtg gttatgtttc catggttacc 3600
atgggtgtgg atggcagtgt ggggagggca ggtccagctc tgtgggccct accctcctgc 3660
tgagctgccc ctgctgctta agtgcatgca ttgagctgcc tccagcctgg tggcccagct 3720
attaccacac ttggggttta aatatccagg tgtgcccctc caagtcagaa agagatgtcc 3780
ttgtaatatt ccctttagg tgagggttgg taaggggttg gtatctcagg tctgaatctt 3840
caccatcttt ctgattccgc accctgccta cgccaggaga agttgagggg agcatgcttc 3900
cctgcagctg accgggtcac acaaaggcat gctggagtac ccagcctatc aggtgccccc 3960
cttccaaagg cagcgtgccg agccagcaag aggaaggggt gctgtgaggc ttgcccagga 4020
gcaagtgagg ccggagagga gttcaggaac ccttctccat acccacaatc tgagcacgct 4080
accaaatctc aaaatatcct aagactaaca aaggcagctg tgtctgagcc caacccttct 4140
aaacggtgac ctttagtgcc aacttccccc ctaactggac agcctcttct gtcccaagtc 4200
tccagagaga aatcaggcct gatgaggggg aattcctgga acctggaccc cagccttggt 4260
gggggagcct ctggaatgca tggggcgggt cctagctgtt agggacattt ccaagctgtt 4320
agttgctgtt taaaatagaa ataaaattga agactaaaga ccta 4364

<210> 14
<211> 882
<212> PRT
<213> Homo sapiens

<400> 14
Met Ala Leu Arg Arg Ser Met Gly Arg Pro Gly Leu Pro Pro Leu Pro
1 5 10 15

Leu Pro Pro Pro Pro Arg Leu Gly Leu Leu Leu Ala Glu Ser Ala Ala
20 25 30

Ala Gly Leu Lys Leu Met Gly Ala Pro Val Lys Leu Thr Val Ser Gln

35 40 45

Gly Gln Pro Val Lys Leu Asn Cys Ser Val Glu Gly Met Glu Glu Pro
50 55 60

Asp Ile Gln Trp Val Lys Asp Gly Ala Val Val Gln Asn Leu Asp Gln
65 70 75 80

Leu Tyr Ile Pro Val Ser Glu Gln His Trp Ile Gly Phe Leu Ser Leu
85 90 95

Lys Ser Val Glu Arg Ser Asp Ala Gly Arg Tyr Trp Cys Gln Val Glu
100 105 110

Asp Gly Gly Glu Thr Glu Ile Ser Gln Pro Val Trp Leu Thr Val Glu
115 120 125

Gly Val Pro Phe Phe Thr Val Glu Pro Lys Asp Leu Ala Val Pro Pro
130 135 140

Asn Ala Pro Phe Gln Leu Ser Cys Glu Ala Val Gly Pro Pro Glu Pro
145 150 155 160

Val Thr Ile Val Trp Trp Arg Gly Thr Thr Lys Ile Gly Gly Pro Ala
165 170 175

Pro Ser Pro Ser Val Leu Asn Val Thr Gly Val Thr Gln Ser Thr Met
180 185 190

Phe Ser Cys Glu Ala His Asn Leu Lys Gly Leu Ala Ser Ser Arg Thr
195 200 205

Ala Thr Val His Leu Gln Ala Leu Pro Ala Ala Pro Phe Asn Ile Thr
210 215 220

Val Thr Lys Leu Ser Ser Ser Asn Ala Ser Val Ala Trp Met Pro Gly
225 230 235 240

Ala Asp Gly Arg Ala Leu Leu Gln Ser Cys Thr Val Gln Val Thr Gln
245 250 255

Ala Pro Gly Gly Trp Glu Val Leu Ala Val Val Val Pro Val Pro Pro
260 265 270

Phe Thr Cys Leu Leu Arg Asp Leu Val Pro Ala Thr Asn Tyr Ser Leu
275 280 285

Arg Val Arg Cys Ala Asn Ala Leu Gly Pro Ser Pro Tyr Ala Asp Trp
290 295 300

Val Pro Phe Gln Thr Lys Gly Leu Ala Pro Ala Ser Ala Pro Gln Asn
305 310 315 320

Leu His Ala Ile Arg Thr Asp Ser Gly Leu Ile Leu Glu Trp Glu Glu
325 330 335

Val Ile Pro Glu Ala Pro Leu Glu Gly Pro Leu Gly Pro Tyr Lys Leu
340 345 350

Ser Trp Val Gln Asp Asn Gly Thr Gln Asp Glu Leu Thr Val Glu Gly
355 360 365

Thr Arg Ala Asn Leu Thr Gly Trp Asp Pro Gln Lys Asp Leu Ile Val
370 375 380

Arg Val Cys Val Ser Asn Ala Val Gly Cys Gly Pro Trp Ser Gln Pro
385 390 395 400

Leu Val Val Ser Ser His Asp Arg Ala Gly Gln Gln Gly Pro Pro His
405 410 415

Ser Arg Thr Ser Trp Val Pro Val Val Leu Gly Val Leu Thr Ala Leu
420 425 430

Val Thr Ala Ala Ala Leu Ala Leu Ile Leu Leu Arg Lys Arg Arg Lys
435 440 445

Glu Thr Arg Phe Gly Gln Ala Phe Asp Ser Val Met Ala Arg Gly Glu
450 455 460

Pro Ala Val His Phe Arg Ala Ala Arg Ser Phe Asn Arg Glu Arg Pro
465 470 475 480

Glu Arg Ile Glu Ala Thr Leu Asp Ser Leu Gly Ile Ser Asp Glu Leu
485 490 495

Lys Glu Lys Leu Glu Asp Val Leu Ile Pro Glu Gln Gln Phe Thr Leu
500 505 510

Gly Arg Met Leu Gly Lys Gly Glu Phe Gly Ser Val Arg Glu Ala Gln
515 520 525

Leu Lys Gln Glu Asp Gly Ser Phe Val Lys Val Ala Val Lys Met Leu
530 535 540

Lys Ala Asp Ile Ile Ala Ser Ser Asp Ile Glu Glu Phe Leu Arg Glu
545 550 555 560

Ala Ala Cys Met Lys Glu Phe Asp His Pro His Val Ala Lys Leu Val
565 570 575

Gly Val Ser Leu Arg Ser Arg Ala Lys Gly Arg Leu Pro Ile Pro Met
580 585 590

Val Ile Leu Pro Phe Met Lys His Gly Asp Leu His Ala Phe Leu Leu
595 600 605

Ala Ser Arg Ile Gly Glu Asn Pro Phe Asn Leu Pro Leu Gln Thr Leu
610 615 620

Ile Arg Phe Met Val Asp Ile Ala Cys Gly Met Glu Tyr Leu Ser Ser
625 630 635 640

Arg Asn Phe Ile His Arg Asp Leu Ala Ala Arg Asn Cys Met Leu Ala
645 650 655

Glu Asp Met Thr Val Cys Val Ala Asp Phe Gly Leu Ser Arg Lys Ile
660 665 670

Tyr Ser Gly Asp Tyr Tyr Arg Gln Gly Cys Ala Ser Lys Leu Pro Val
675 680 685

Lys Trp Leu Ala Leu Glu Ser Leu Ala Asp Asn Leu Tyr Thr Val Gln
690 695 700

Ser Asp Val Trp Ala Phe Gly Val Thr Met Trp Glu Ile Met Thr Arg
705 710 715 720

Gly Gln Thr Pro Tyr Ala Gly Ile Glu Asn Ala Glu Ile Tyr Asn Tyr
725 730 735

Leu Ile Gly Gly Asn Arg Leu Lys Gln Pro Pro Glu Cys Met Glu Asp
740 745 750

Val Tyr Asp Leu Met Tyr Gln Cys Trp Ser Ala Asp Pro Lys Gln Arg
755 760 765

Pro Ser Phe Thr Cys Leu Arg Met Glu Leu Glu Asn Ile Leu Gly Gln
770 775 780

Leu Ser Val Leu Ser Ala Ser Gln Asp Pro Leu Tyr Ile Asn Ile Glu
785 790 795 800

Arg Ala Glu Glu Pro Thr Val Gly Gly Ser Leu Glu Leu Pro Gly Arg
805 810 815

Asp Gln Pro Tyr Ser Gly Ala Gly Asp Gly Ser Gly Met Gly Ala Val
820 825 830

Gly Gly Thr Pro Ser Asp Cys Arg Tyr Ile Leu Thr Pro Gly Gly Leu
835 840 845

Ala Glu Gln Pro Gly Gln Ala Glu His Gln Pro Glu Ser Pro Leu Asn
850 855 860

Glu Thr Gln Arg Leu Leu Leu Leu Gln Gln Gly Leu Leu Pro His Ser
865 870 875 880

Ser Cys

<210> 15
<211> 1457
<212> DNA
<213> Homo sapiens

<400> 15
ggccgcttcg ggtttccgga ggggccggag ggcgggcgag ggcgtcacgt gcgcgccgcc 60
cgcgggccgg ttggtccccg ggcgggggag gggccgtgcg cagcctgggt cggggtcggg 120
ccggggtcgg cacctgggac atccctgagg gaagggccgg gagcgggagc gccccagcgg 180
ccggcgggcg ggcgggcgag cggacgagcg gcgcggagcc ggcccgaggc gcgcgccgag 240
ggagccccgt ccccggtcgt gggggcaccg cccgcaggct ctgcggggtg ggcagctccc 300
gggcctgcca tgagctctcc gccgcccgcc cgcagtggct tttaccgcca ggaggtgacc 360
aagacggcct gggaggtgcg cgccgtgtac cgggacctgc agcccgtggg ctcgggcgcc 420
tacggcgcgg tgtgctcggc cgtggacggc cgcaccggcg ctaaggtggc catcaagaag 480
ctgtatcggc ccttccagtc cgagctgttc gccaagcgcg cctaccgcga gctgcgcctg 540
ctcaagcaca tgcgccacga gaacgtgatc gggctgctgg acgtattcac tcctgatgag 600

```
accctggatg acttcacgga cttttacctg gtgatgccgt tcatgggcac cgacctgggc    660
aagctcatga aacatgagaa gctaggcgag gaccggatcc agttcctcgt gtaccagatg    720
ctgaaggggc tgaggtatat ccacgctgcc ggcatcatcc acagagacct gaagcccggc    780
aacctggctg tgaacgaaga ctgtgagctg aagatcctgg acttcggcct ggccaggcag    840
gcagacagtg agatgactgg gtacgtggtg acccggtggt accgggctcc cgaggtcatc    900
ttgaattgga tgcgctacac gcagacggtg gacatctggt ccgtgggctg catcatggcg    960
gagatgatca caggcaagac gctgttcaag ggcagcgacc acctggacca gctgaaggag   1020
atcatgaagg tgacggggac gcctccggct gagtttgtgc agcggctgca gagcgatgag   1080
gccaagaaca acatgaaggg cctccccgaa ttggagaaga aggattttgc ctctatcctg   1140
accaatgcaa gccctctggc tgtgaacctc ctggagaaga tgctggtgct ggacgcggag   1200
cagcgggtga cggcaggcga ggcgctggcc catccctact tcgagtccct gcacgacacg   1260
gaagatgagc cccaggtcca gaagtatgat gactcctttg acgacgttga ccgcacactg   1320
gatgaatgga agcgtgttac ttacaaagag gtgctcagct tcaagcctcc ccggcagctg   1380
ggggccaggg tctccaagga gacgcctctg tgaagatctc tgggctccgg ggtggcagtg   1440
aggaccacct tcacctt                                                  1457
```

<210> 16
<211> 367
<212> PRT
<213> Homo sapiens

<400> 16

```
Met Ser Ser Pro Pro Pro Ala Arg Ser Gly Phe Tyr Arg Gln Glu Val
1               5                  10                  15

Thr Lys Thr Ala Trp Glu Val Arg Ala Val Tyr Arg Asp Leu Gln Pro
            20                  25                  30

Val Gly Ser Gly Ala Tyr Gly Ala Val Cys Ser Ala Val Asp Gly Arg
        35                  40                  45

Thr Gly Ala Lys Val Ala Ile Lys Lys Leu Tyr Arg Pro Phe Gln Ser
    50                  55                  60

Glu Leu Phe Ala Lys Arg Ala Tyr Arg Glu Leu Arg Leu Leu Lys His
65                  70                  75                  80

Met Arg His Glu Asn Val Ile Gly Leu Leu Asp Val Phe Thr Pro Asp
                85                  90                  95

Glu Thr Leu Asp Asp Phe Thr Asp Phe Tyr Leu Val Met Pro Phe Met
            100                 105                 110

Gly Thr Asp Leu Gly Lys Leu Met Lys His Glu Lys Leu Gly Glu Asp
        115                 120                 125

Arg Ile Gln Phe Leu Val Tyr Gln Met Leu Lys Gly Leu Arg Tyr Ile
    130                 135                 140

His Ala Ala Gly Ile Ile His Arg Asp Leu Lys Pro Gly Asn Leu Ala
145                 150                 155                 160

Val Asn Glu Asp Cys Glu Leu Lys Ile Leu Asp Phe Gly Leu Ala Arg
                165                 170                 175

Gln Ala Asp Ser Glu Met Thr Gly Tyr Val Val Thr Arg Trp Tyr Arg
            180                 185                 190
```

```
Ala Pro Glu Val Ile Leu Asn Trp Met Arg Tyr Thr Gln Thr Val Asp
        195                 200                 205

Ile Trp Ser Val Gly Cys Ile Met Ala Glu Met Ile Thr Gly Lys Thr
    210                 215                 220

Leu Phe Lys Gly Ser Asp His Leu Asp Gln Leu Lys Glu Ile Met Lys
225                 230                 235                 240

Val Thr Gly Thr Pro Pro Ala Glu Phe Val Gln Arg Leu Gln Ser Asp
                245                 250                 255

Glu Ala Lys Asn Asn Met Lys Gly Leu Pro Glu Leu Glu Lys Lys Asp
            260                 265                 270

Phe Ala Ser Ile Leu Thr Asn Ala Ser Pro Leu Ala Val Asn Leu Leu
        275                 280                 285

Glu Lys Met Leu Val Leu Asp Ala Glu Gln Arg Val Thr Ala Gly Glu
    290                 295                 300

Ala Leu Ala His Pro Tyr Phe Glu Ser Leu His Asp Thr Glu Asp Glu
305                 310                 315                 320

Pro Gln Val Gln Lys Tyr Asp Asp Ser Phe Asp Asp Val Asp Arg Thr
                325                 330                 335

Leu Asp Glu Trp Lys Arg Val Thr Tyr Lys Glu Val Leu Ser Phe Lys
            340                 345                 350

Pro Pro Arg Gln Leu Gly Ala Arg Val Ser Lys Glu Thr Pro Leu
        355                 360                 365
```

<210> 17
<211> 5243
<212> DNA
<213> Homo sapiens

<400> 17

```
cttttcttgc aggacatgtt ctctggatgt cagctgagtc attaaagtaa ctctgcatgt     60
cagtagacag accttggtag aaccacaagg ctcccagaga cacccatctc tcctcatttt    120
tttggtgtgt gtgtcttcac cgaacattca aaactgtttc tccaaagcgt tttgcaaaaa    180
ctcagactgt tttccaaagc agaagcactg gagtccccag cagaagcgat gggcagtgtg    240
cgaaccaacc gctacagcat cgtctcttca gaagaagacg gtatgaagtt ggccaccatg    300
gcagttgcaa atggctttgg gaacgggaag agtaaagtcc acacccgaca acagtgcagg    360
agccgctttg tgaagaaaga tggccactgt aatgttcagt tcatcaatgt gggtgagaag    420
gggcaacggt acctcgcaga catcttcacc acgtgtgtgg acattcgctg gcggtggatg    480
ctggttatct tctgcctggc tttcgtcctg tcatggctgt tttttggctg tgtgttttgg    540
ttgatagctc tgctccatgg ggacctggat gcatccaaag agggcaaagc ttgtgtgtcc    600
gaggtcaaca gcttcacggc tgccttcctc ttctccattg agacccagac aaccataggc    660
tatggtttca gatgtgtcac ggatgaatgc ccaattgctg ttttcatggt ggtgttccag    720
tcaatcgtgg gctgcatcat cgatgctttc atcattggcg cagtcatggc caagatggca    780
aagccaaaga agagaaacga gactcttgtc ttcagtcaca atgccgtgat tgccatgaga    840
gacggcaagc tgtgtttgat gtggcgagtg ggcaatcttc ggaaaagcca cttggtggaa    900
gctcatgttc gagcacagct cctcaaatcc agaattactt ctgaagggga gtatatccct    960
ctggatcaaa tagacatcaa tgttgggttt gacagtggaa tcgatcgtat atttctggtg   1020
tccccaatca ctatagtcca tgaaatagat gaagacagtc ctttatatga tttgagtaaa   1080
caggacattg acaacgcaga ctttgaaatc gtggtcatac tggaaggcat ggtggaagcc   1140
actgccatga cgacacagtg ccgtagctct tatctagcaa atgaaatcct gtggggccac   1200
```

cgctatgagc ctgtgctctt tgaagagaag cactactaca aagtggacta ttccaggttc 1260
cacaaaactt acgaagtccc caacactccc ctttgtagtg ccagagactt agcagaaaag 1320
aaatatatcc tctcaaatgc aaattcattt tgctatgaaa atgaagttgc cctcacaagc 1380
aaagaggaag acgacagtga aaatggagtt ccagaaagca ctagtacgga cacgcccccc 1440
gacatagacc ttcacaacca ggcaagtgta cctctagagc ccaggccctt acggcgagag 1500
tcggagatat gactgactga ttccttctct ggaatagtta ctttacaaca cggtctgttg 1560
gtcagaggcc caaaacagtt atacagatga cggtactggt caagatgggt caagcaagcg 1620
gccacaaggg actgaggcaa gcacaatggt ttcaaagaaa gactgtaagc tccatgatta 1680
gcataaagca ctaaccatgt ctccatgtga cccgatggca catagatgtt gtagaataag 1740
ttatgggttt ttatgttttg ttttgtgttt ttccaaaact tgaacttgca ggcaagcctt 1800
ggttgggtat ttgatttatc cagaatgctt ctctttaggg aacaaggatg tttttaatgg 1860
cataacaaag gcaagactct gccttaattt ttgaaaagct gctaactaca tgaacacaaa 1920
tgtgtatttt tgttgcagtg tagttttcct tttgtgtaat tttaaagtca gtgttgaatt 1980
ttattgaaag ctcatgatgc gcttcaaagt ggcaagtatt tggctattaa ctgccaaaac 2040
aagagcctga ttttttgagg ccagtaattc gtttgctaga attgattttt tttctctctc 2100
tctttgttac ataagggcat tatgtaacac tagccgaatg gtagcctctg ggttgttgtt 2160
ttttttcttt cctccatgat gttaatgggt tatctcaaat tttaagttaa actacctaaa 2220
ataaatacca aagataatgc atatttttgc acagtggagc ttacacttaa aagaaaacaa 2280
agccccatgg gctgccttga aatcaagaga caataacttt gaacctcagc aagaccttga 2340
accgccggtt cattttgcac cttattcaga aaatagagca tcatactcac cgagtctagt 2400
cagtgtagtg cttttaaaaa ttttgtcctt tcatgtaact tttttatttt aagaggaaga 2460
agaagaaagg ggcacacaca cacaataccg acgtctatcc tttcctgcta ggcagtgctg 2520
gccaggctca tgtgtagtgt gcgagatggt gatgtactct tatatttttc tgggcttttc 2580
cttttgcaca ttccaaaatt catttcataa gacaagatct tcataggacc tccttggcat 2640
cctggcattc tcaaaactga gccatccagc atgaaagata aatgggttta aacccttgct 2700
gctgaattta ttgcctggac tgtcaggaca tcaccagccc accttcacct tagggaagat 2760
gccacacctg gcctccacac ttgctcttct gatcagtctg tctggattga gtcctacagt 2820
gtcagatagg gcggcaaatg ccaaagcagg gaaacaggga ggtgtggaca agccagtttg 2880
atgcagcact tcagatcaag tgcttaggaa ggagaggaaa cttgcctttt ttatggcaga 2940
ggatagtaat gaaaatgtct cagtatttta gggtcaatga gagccataaa aatataacat 3000
aatcacaagt aaaggagata atggtctaaa acagctattt ccctttttctg tgtgcatact 3060
tatgactgaa tgtgagctaa gcattttctc ctgtggagcc ctagagcagg ttactaagga 3120
aggacacatt gttttccaga agcctcccct gcctggctga ctgccttgct agaaacataa 3180
ttttttttt ctcactgaag ctcaataatg gaactctttt ttttttttt tttaatttaa 3240
agttccctat ttgtgaattc tgggattact gacttttctt tttaattgga gtctcaaaat 3300
caactctctt atggtattat atctctgtat gccattaaaa aacagcttgt tctagaatca 3360
tgtattttgt aaactgatgt ttgtgatggt ctctggttct tgaacagcca tatctgaatg 3420
ccgtgcctgc aaaactatga caatttttgc tgttttcagc cttcagattt gatggcttgg 3480
gaaactgagg tgttattttc aatgaaacaa agaaagagat gttaagcaag tggttgtttt 3540
agatccaaat gtaaaggcag gtttgggaag gtgtttaaag agttggagga attggggatt 3600
gagttgtaaa gaaaacttac agaagaggca acaatttggt tcttgacagt gagaggatat 3660
tgagggcttc agctgctgct attatgatgt tttgcaaagg aaaataatca aaccaaagag 3720
tattcagtga tatgtaaatt aaatgaagat acagtggaga atgggggtga ccacaaaaga 3780
ggctcccccc aaacacacag tgctgccact taaaaagact tgagaaattt gaaaggggct 3840
gggtatgggg ggggcaagaa agagggaggg aaatctttca acttatttct gaaaaagaga 3900
aaaaaatata aaatttctgg tgcacaggtt tgtttttttca agaaaatttt gcagaagcta 3960
tgtttttaaa gtgtacattt tataaagttt atcagatatt ttcatattta aagccaaatg 4020
taaatagagg tctgtaaaga aaaataattg ccatagaaag tataatttca gtgcagtaat 4080
ttctgagagc tagtacctat atgctaccgg ttagcatggt tttagcaaat atataccagc 4140
cttataaggt tcgtattgct atgttcttct gttatttatt tcagcatgga ctgttcattt 4200
gaaacctttt tctagttatt agcgttttaa cagttacaag ctttaaatgg caatttttct 4260
ttttttttct ttttttttct ttttttgtc aagagccaag acacaggtaa tgcacgacat 4320
tgattgctgc attttacctt caaaatattt gtccttattg actgggtctc cttaattaat 4380
gtacacatgt cattagaatg cagacggagg ggactcacca tgaatatctg gggttgattc 4440
ccagatgtgt gttgcttctc tattgcaagc agattccctg ttggatttac ttcggattta 4500
ttccctttta aagaattttt gcccatatct ggaagggcac tatatttttg ggaggagcca 4560
tagattcctg gttatcctat ttttaaacaa aatgtagaca aagtgaactc tattttgatt 4620
attgagaaag gagtagtttt ctatccctct aagagtatac ttgaatcaga cattttaagg 4680
atgtcactat ggcactgttg tcatttccaa attcctagaa aagtttgttt tactttgttt 4740

```
ttattctgtt aatgcattct ttcttctctt tacttccttt cttaccagta cactcctatc   4800
tcaactctgt ttatttgatg agttctgtcc cgtaaatcat atttcccttа caattaataa   4860
atgtcacttc atattttata ataaaccact cagtaaaagc aaaagcttgt cctgagaagt   4920
agagtgagtt ctttttcact ctgtgtctaa taatgttaag gtgggaaaaa aaaaagtgtg   4980
gcatagctac ctgcccatcc ccaaccctca gcaaagtaga atctcttttc tggtaatttt   5040
gggtttccgc tctgggctct ggcaagttga acaatcctag ccattgacaa tcgtgatagt   5100
tattattttc ccatttgctg tctttttgta tctaaagtct tcctattgta ctgcacaaac   5160
catggattgt acatattttt atatattatg tcttatttta ttatttctaa ataaaaaaat   5220
taaaaattga aaacaaattc ttg                                          5243
```

<210> 18
<211> 427
<212> PRT
<213> Homo sapiens

<400> 18

```
Met Gly Ser Val Arg Thr Asn Arg Tyr Ser Ile Val Ser Ser Glu Glu
1               5                  10                  15

Asp Gly Met Lys Leu Ala Thr Met Ala Val Ala Asn Gly Phe Gly Asn
            20                  25                  30

Gly Lys Ser Lys Val His Thr Arg Gln Gln Cys Arg Ser Arg Phe Val
        35                  40                  45

Lys Lys Asp Gly His Cys Asn Val Gln Phe Ile Asn Val Gly Glu Lys
    50                  55                  60

Gly Gln Arg Tyr Leu Ala Asp Ile Phe Thr Thr Cys Val Asp Ile Arg
65                  70                  75                  80

Trp Arg Trp Met Leu Val Ile Phe Cys Leu Ala Phe Val Leu Ser Trp
                85                  90                  95

Leu Phe Phe Gly Cys Val Phe Trp Leu Ile Ala Leu Leu His Gly Asp
            100                 105                 110

Leu Asp Ala Ser Lys Glu Gly Lys Ala Cys Val Ser Glu Val Asn Ser
        115                 120                 125

Phe Thr Ala Ala Phe Leu Phe Ser Ile Glu Thr Gln Thr Thr Ile Gly
    130                 135                 140

Tyr Gly Phe Arg Cys Val Thr Asp Glu Cys Pro Ile Ala Val Phe Met
145                 150                 155                 160

Val Val Phe Gln Ser Ile Val Gly Cys Ile Ile Asp Ala Phe Ile Ile
                165                 170                 175

Gly Ala Val Met Ala Lys Met Ala Lys Pro Lys Lys Arg Asn Glu Thr
            180                 185                 190

Leu Val Phe Ser His Asn Ala Val Ile Ala Met Arg Asp Gly Lys Leu
        195                 200                 205

Cys Leu Met Trp Arg Val Gly Asn Leu Arg Lys Ser His Leu Val Glu
    210                 215                 220
```

Ala His Val Arg Ala Gln Leu Leu Lys Ser Arg Ile Thr Ser Glu Gly
225 230 235 240

Glu Tyr Ile Pro Leu Asp Gln Ile Asp Ile Asn Val Gly Phe Asp Ser
245 250 255

Gly Ile Asp Arg Ile Phe Leu Val Ser Pro Ile Thr Ile Val His Glu
260 265 270

Ile Asp Glu Asp Ser Pro Leu Tyr Asp Leu Ser Lys Gln Asp Ile Asp
275 280 285

Asn Ala Asp Phe Glu Ile Val Val Ile Leu Glu Gly Met Val Glu Ala
290 295 300

Thr Ala Met Thr Thr Gln Cys Arg Ser Ser Tyr Leu Ala Asn Glu Ile
305 310 315 320

Leu Trp Gly His Arg Tyr Glu Pro Val Leu Phe Glu Glu Lys His Tyr
325 330 335

Tyr Lys Val Asp Tyr Ser Arg Phe His Lys Thr Tyr Glu Val Pro Asn
340 345 350

Thr Pro Leu Cys Ser Ala Arg Asp Leu Ala Glu Lys Lys Tyr Ile Leu
355 360 365

Ser Asn Ala Asn Ser Phe Cys Tyr Glu Asn Glu Val Ala Leu Thr Ser
370 375 380

Lys Glu Glu Asp Asp Ser Glu Asn Gly Val Pro Glu Ser Thr Ser Thr
385 390 395 400

Asp Thr Pro Pro Asp Ile Asp Leu His Asn Gln Ala Ser Val Pro Leu
405 410 415

Glu Pro Arg Pro Leu Arg Arg Glu Ser Glu Ile
420 425

<210> 19
<211> 1050
<212> DNA
<213> Homo sapiens

<400> 19
tttatatttt ttcagtgtcc atatttcaaa aatttattta tctcaaactg tgcataatgg 60
agtaaaaact taagttgaaa atgtacctgt tataaggatg atattagttc aaatatatac 120
atggattctc ggcagactga ttcaaataat acagagccga atcttttaaa atacaactac 180
ggaaaataaa ggggggaaaa ccttaaaata tcacaataaa tttacagaaa tattacaaac 240
cataagaaaa tatttcaaac acagtaattt catggttttt tttatctgaa caaaatggaa 300
agttgggatc gaacaaaagc tattataaat taccaacggt gtcaacctgc atggccattt 360
ttgcttttaa cagtaagtta taaaatttag tacagtctaa aacttttgcc ctttttaaac 420
aagaccacag agatggttcg ccagtactta ttctaatttt ttccttttgt acaattttta 480
aacaattaaa atgtccaaat ttgaataatt ttcttctttt cacgtttgca actgtcctaa 540
atttcagctg cagaatcaaa attcagcaag aagcctctcc ttgaaaaata ttggcaaatt 600
ctcagcttat aaacaatgga cattttgatt gccatgttta tctcgataaa tactgtacaa 660
aagttgcttg caaatattaa aacatttttt tcgtcgcttg gagactagct ctaaatatta 720
ttggtaaaga cttttgcaaa cttcctgcaa agctcctacc gtaccactag aacttttaaa 780
aagtttttcg tagctttctt tcctccagat ctatacaagg tccattcccc cgccctcccc 840

accctcccca ggttttctct gtacaaaaat agtcccccaa aaagaagtcc aggatctctc 900
tcataaaagt tttcttgtcg gcatcgcggt ttttgcgtga gtgtggatgg gattggtgtt 960
ctcttttgca gctgtcattt gctgtgggtg atgggatttt ttttttcct ttttcttttt 1020
gagcgtaccg ggttttctcc atgctgtttc 1050

<210> 20
<211> 88
<212> PRT
<213> Homo sapiens

<400> 20
Met Glu Lys Thr Arg Tyr Ala Gln Lys Glu Lys Gly Lys Lys Lys Ile
1 5 10 15

Pro Ser Pro Thr Ala Asn Asp Ser Cys Lys Arg Glu His Gln Ser His
20 25 30

Pro His Ser Arg Lys Asn Arg Asp Ala Asp Lys Lys Thr Phe Met Arg
35 40 45

Glu Ile Leu Asp Phe Phe Leu Gly Asp Tyr Phe Cys Thr Glu Lys Thr
50 55 60

Trp Gly Gly Trp Gly Gly Arg Gly Asn Gly Pro Cys Ile Asp Leu Glu
65 70 75 80

Glu Arg Lys Leu Arg Lys Thr Phe
85

<210> 21
<211> 881
<212> DNA
<213> Homo sapiens

<400> 21
ttttttttt ttgtcaagag ccaagacaca ggtaatgcac gacattgatt gctgcatttt 60
accttcaaaa tatttgtcct tattgactgg gtctccttaa ttaatgtaca catgtcatta 120
gaatgcagac ggaggggact caccatgaat attctgggtt gattcccaga tgtgtgttgc 180
ttctctattg caagcagatt ccctgttgga tttacttcgg atttattccc ttttaaagaa 240
tttttgccca tatctggaag ggcactatat ttttgggagg agccatagat tcctggttat 300
cctatttta aacaaaatgt agacaaagtg aactctattt tgattattga gaaaggagta 360
gttttctatc cctctaagag tatacttgaa tcagacattt taaggatgtc actatggcac 420
tgttgtcatt tccaaattcc tagaaaagtt tgttttactt tgtttttatt ctgttaatgc 480
attctttctt ctctttactt cctttcttac cagtacactc ctatctcaac tctgtttatt 540
tgatgagttc tgtcccgtaa atcatatttc ccttacaatt aataaatgtc acttcatatt 600
ttataataaa ccactcagta aaagcaaaag cttgtcctga gaagtactct gtgtctaata 660
atgttaaggg catagctacc tgcccatccc caaccctcag caaagtagaa tctcttttct 720
ggtaattttg ggtttccgct ctgggctctg gcaagttgaa caatcctagc cattgacaat 780
cgtgatagtt attattttcc catttgctgt ctttttgtat ctaaagtctt cctattgtac 840
tgcacaaacc atggattgta catattttta tatattatgt c 881